

MEANING AND PROSODY:
ON THE WEB, IN THE LAB AND FROM THE THEORIST'S ARMCHAIR

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Jonathan Anthony Howell

January 2012

© 2012 Jonathan Anthony Howell

MEANING AND PROSODY:
ON THE WEB, IN THE LAB AND FROM THE THEORIST'S ARMCHAIR

Jonathan Anthony Howell, Ph. D.

Cornell University 2012

I present a new approach to research on meaning and prosody, using speech “harvested” from the web. I advocate a pluralistic view of linguistic data and methodology, within which web-harvested speech plays a vital role. I show that web-harvested speech can be used effectively with computational and experimental methods on the one hand, and qualitative, impressionistic study on the other.

My domain of inquiry is the well-known correlation between
(i) which information in a discourse is most important (e.g. new or contrastive); and
(ii) which material in an utterance is realized with prosodic prominence (e.g. stress, accent)
which I refer to as “focus”.

In Chapter 2, I describe the method of harvesting speech data from the web, quantify its efficacy and discuss possible improvements.

In Chapter 3, I investigate the location and acoustic realization of focus in comparative clauses (e.g. *than I did*). Using machine learning and human classifiers, I discover a robust correlation between particular acoustic cues of prosodic prominence and the location of focus predicted by linguistic theory. From the robustness of non-intonational acoustic cues, I hypothesize that focus may be realized by discrete, paradigmatic (i.e. cross-utterance) categories of stress. Results obtained from the

web-harvested speech are cross-validated in a laboratory production experiment with stimuli modeled on the web data. Experimental results also confirm a distinct, but ambiguous prosodic realization of “second occurrence focus”, which has been central to debates surrounding the semantics of focus.

In Chapters 4 and 5, I investigate the adnominal emphatic reflexive (ER; e.g. *himself* in *Jane met Chomsky himself*). I argue that it is an instance of a theoretically predicted but poorly attested focus-sensitive operator having sub-propositional scope. Using constructed data and personal introspection, I argue that the adnominal ER exhibits the expected pragmatic, semantic, syntactic and prosodic properties of focus sensitive constructions, and I reconcile opposing approaches to its semantics. Finally, I debunk a deterministic view of focus, according to which certain linguistic constructions in a language are inherently or obligatorily focused, through the careful investigation of the intonation and discourse context of individual examples of the adnominal ER.

BIOGRAPHICAL SKETCH

Jonathan Howell received a B.A. (Honours) in Linguistics from the University of British Columbia in 2003.

For Catherine

utterance sputterance intonation
bring me acoustic realization
while NASA looks for a new supernova
chomsky chomps on subcorpora
himself herself doing well
but second occurrence focus is a living hell
adnominal emphatic reflexives
packaged as armchair laxatives
web-harvested *and* speech corpora
sounds like a flock of crows flying over
linear discriminant analysis
who even knows what the f__ that is?
in the absence of meaning we need support
vector machines to sort it all out
amidst this cacaphony, all this noise
what a relief -- Dr. Howell's in the house!

-- Rick Stapleton

ACKNOWLEDGMENTS

You will find that many dissertation writers liken the process to a journey. I find this analogy particularly vivid, as I have spent many (many!) hours traveling since I first entered graduate school, most notably in my little white 1990 Honda Accord (with B.C. plates), border-hopping between Ithaca, NY and my wife's various Ontario addresses. Although often tedious, the experience afforded me lots of time for thoughtful contemplation and an appreciation for the culinary delights of the New York State Thruway. I would like to thank the individuals who guided me in my dissertation journey and who encouraged me to always get back in the car (in both directions).

First, I offer sincere gratitude to the members of my dissertation committee. I remember from my first year that Mats Rooth always had the most pointed (and often devastating) remarks at department colloquia, regardless of the topic. Yet even as a beginning student, I always felt comfortable approaching him with the sketchiest of proposals. The breadth and depth of his intellectual interests and expertise have proved invaluable, not to mention his patience at my initial ignorance of the command-line. It was also my great fortune that freshly-minted Ph.D. Michael Wagner arrived at Cornell not long after I did. Adeptly navigating the prosodic and the grammatical, the theoretical and the experimental, he has been both a role model and mentor. He also navigated us in the Honda Accord on route to one of the many conferences at which he encouraged me to present. Abby Cohn generously agreed to join my committee halfway into the venture; every time we met she managed to provide the perfect tonic of insight, advice and encouragement.

I learned a great deal about the phonology and syntax of focus from Molly Diesing and Draga Zec, as well as how to have lots of fun teaching linguistics. Thank you also to Amanda Miller for insisting that I persevere with scripting in *Praat* and for many hours spent debugging my efforts together. I also benefitted from the teaching and advice of Dorit Abusch, John Bowers, Sue Hertz, Sally McConnell-Ginet and John Whitman.

The Honda Accord was often parked outside the apartments of Adam Cooper, Effi Georgala, Johanna Brugman, Nikola Predolac and Esra Kesici. I thank them for their friendship and advice, and for graciously hosting me, sometimes on very short notice. Thank you to all my fellow graduate students, who made graduate school such a positive experience, including Christina Bjorndahl, Marc Brunelle, Hongyuan Dong, Tova Friedman, Masayuki Gibson, Hyun Kyung Hwang, Pittayawat (Joe) Pittayaporn and Jiwon Yun.

Thank you to Michael Wagner, members of the Prosody Lab and the department of linguistics at McGill University for hosting me as a visiting student and providing me with the resources to carry out the laboratory experiments in this dissertation. Funding for this research came from a project “Harvesting Speech Datasets for Linguistic Research on the Web” (principal investigators Mats Rooth and Michael Wagner) jointly funded by the Social Sciences and Humanities Research Council (SSHRC) in Canada and the National Science Foundation (NSF) in the US as part of an initiative called the “Digging into Data Challenge”. (This time I took the train.)

Ivona Kučerová and Tae-Jin Yoon at McMaster University and Keren Rice at the University of Toronto shared their time generously. Particular thanks to Ivona for a timely kick in the pants.

The brilliant and fun-loving faculty at the University of British Columbia first excited my interest in linguistics. A special thank you to Lisa Matthewson for an exuberant and inspiring introduction to semantics and to Bryan Gick for setting me loose in the speech lab.

To my wife's family, I know you're still not sure exactly what I was doing in New York State all that time; thank you for cheering me along anyway. To my own family, your unconditional love and support made this achievement possible. Mom, you taught me how to love language; Dad, you taught me how to have fun with it.

Most especially, to my amazing wife Catherine, thank you for putting up with a 1990 Honda Accord, and a commuting partner, for so many years. I could not have done this without you. And last but not least, thank you to Adelyn, whose imminent arrival precipitated the selling of the Honda Accord. A new journey begins.

TABLE OF CONTENTS

Biographical sketch.....	iii
Dedication.....	iv
Acknowledgments.....	vi
Table of Contents.....	ix
List of Figures.....	xiv
List of Tables.....	xviii
Chapter 1: Introduction.....	1
1 Overview of the dissertation.....	1
2 Objects of study	4
2.1 Correlates of prosodic prominence.....	4
2.2 Correlates of discourse importance.....	16
2.3 Syntactic mediation.....	28
2.4 Domains of focus.....	29
3 Methodological pluralism.....	30
4 The value of web-harvested speech.....	34
4.1 Limits of creativity: non-discovery, false discovery.....	34
4.2 Data quantity, data quality	39
Chapter 2: Web harvest method.....	42
1 Introduction.....	42
2 Web harvest method.....	43
3 Evaluation of retrieval efficacy.....	47
4 Discussion.....	51
Chapter 3: Acoustic classification of focus in a web corpus of comparatives.....	54
1 Introduction.....	54

2 Web-harvested data.....	61
2.1 Web-harvesting.....	61
2.2 Acoustic analysis.....	61
2.3 Feature grouping: duration, quality, quantity	63
3 Laboratory data.....	63
3.1 Stimuli.....	63
3.2 Recording.....	64
3.3 Acoustic analysis.....	64
4 Statistical classifiers and feature selection.....	65
4.1 Support vector machines and linear discriminant analysis.....	65
4.2 Redundant features and feature selection.....	70
5 Machine learning experiments.....	74
5.1 Datasets and evaluation of classifier performance.....	74
5.2 Web-trained machine learning classifiers.....	77
5.2.1 Feature selection.....	77
5.2.2 Classification accuracy on web-harvested data.....	84
5.2.3 Classification accuracy on laboratory-elicited data.....	85
5.2.4 Declaratives and interrogatives.....	97
5.2.5 Second occurrence focus.....	100
5.2.6 Discussion.....	101
5.3 Training with laboratory data.....	103
5.3.1 Feature selection.....	103
5.3.2 Classification accuracy on web-harvested data.....	105
5.3.3 Second occurrence focus.....	105
5.3.4 Discussion.....	106

6 Human acoustic classifiers.....	107
6.1 Experiment 1: web stimuli.....	108
6.1.1 Methodology.....	108
6.1.2 Results.....	109
6.1.3 Discussion.....	115
6.2 Experiment 2: laboratory production stimuli.....	116
6.2.1 Methodology.....	116
6.2.2 Results.....	117
6.2.3 Discussion.....	127
7 Conclusion.....	129
7.1 Discussion of results.....	129
7.2 Future directions.....	138
Chapter 4: Focus sensitive operators below the proposition: the case of the adnominal emphatic reflexive.....	141
1 Introduction.....	141
2 Preliminaries.....	143
2.1 What adnominal emphatic reflexives aren't.....	143
2.2 Syntactic distribution.....	151
2.3 Constituency.....	154
2.4 Agreement and c-command.....	155
2.5 Previous formal semantic accounts.....	156
3 Properties of focus sensitive operators.....	158
3.1 A focus sensitive operator c-commands its associate.....	158
3.2 A focus associate has a focus exponent.....	164
3.2.1 Qualitative evidence.....	165

3.2.2 Quantitative evidence.....	165
3.2.3 Prominence in a domain.....	168
3.3 Focus operators operate on propositions only.....	168
3.3.1 The adnominal ER ranks individuals, not propositions.....	172
3.3.2 The adnominal ER within an alternative semantics for focus.....	178
3.3.3 The adnominal ER within a covert movement analysis of focus.....	181
3.4 A focus sensitive operator has distinct focus sensitive interpretations.....	186
3.4.1 Association within the DP.....	186
3.4.2 Scalarity without focus.....	188
3.5 Focus operators may interact.....	190
4 The extended focus sensitive operator approach.....	191
4.1 Reconciling approaches to the adnominal ER.....	191
4.2 Alternatives and discourse antecedence.....	192
5 Conclusion.....	201
Chapter 5: Debunking focal determinism: the case of adnominal emphatic reflexives.....	
1 Introduction.....	203
2 Clausal focus.....	210
3 Argument focus.....	220
4 Focus on the adnominal ER.....	226
5 Focus below the clause.....	234

6 Double focus.....	242
6.1 Double opposition approach.....	245
6.2 Prosody of the double focus configuration.....	249
6.3 Web examples of the double focus configuration.....	254
7 Contrastive topic.....	263
8 Predicate focus and overlapping associate focus.....	267
9 Conclusion.....	270
Appendix A: Acoustic measures.....	275
Appendix B: Stimuli for perception experiment.....	278
References.....	279

LIST OF FIGURES

Chapter 1

Figure 1: Example waveform and f0 contour for an utterance of “I implied that you”	6
Figure 2: <i>F0</i> contour with a series of downstepped H* pitch accents from Liberman & Pierrehumbert (1984).....	15
Figure 3: Smoothed f0 track of (45). Clausal focus configuration.....	37

Chapter 2

Figure 1: Workflow for mp3 retrieval and editing.....	45
Figure 2: Detailed retrieval efficacy at different processing stages compared for 4 different retrieval runs.....	49
Figure 3: Comparison for each target expression of (a) number of tokens in the Switchboard corpus, (b) number of good tokens already collected and identified in the web-harvested corpus and (c) the number of projected tokens available through Everyzing at the time of harvest, based on total hit count and assuming 50% retrieval efficacy.....	50

Chapter 3

Figure 1: Manual phonetic annotation of a token of “than I did.....	62
Figure 2: Separating hyperplane in two-dimensional space (based on figure in Cristianini & Shawe-Taylor 2000).....	66
Figure 3: Linear discriminant analysis (LDA) classifier looks for the optimal model which minimizes within-class distance and maximizes between-class distance.....	68
Figure 4: Optimal separating hyperplane in a two-dimensional space (reproduced from Abe 2005:17).....	69

Figure 5: Scatterplot matrix of four duration-related measures for <i>web1</i>	79
Figure 6: Scatterplot matrix of four quantity-related measures for <i>web1</i>	80
Figure 7: Scatterplot matrix of four quality-related measures for <i>web1</i>	81
Figure 8: Scatterplot matrix of four best measures chosen automatically using the VarSelRF algorithm for <i>web1</i>	82
Figure 9: Scatterplot matrix of four experimenter-selected measures (‘Hand-picked C’) for <i>web1</i>	83
Figure 10: Scatterplot matrix of four duration measures for <i>lab_htk_FOF</i>	88
Figure 11: Scatterplot matrix of four duration measures for <i>lab_hand_FOF</i>	89
Figure 12: Scatterplot matrix for four quantity-related measures for <i>lab_htk_FOF</i>	91
Figure 13: Scatterplot matrix for four quantity-related measures for <i>lab_hand_FOF</i>	92
Figure 14: Scatterplot matrix of four quality-related measures for <i>lab_htk_FOF</i>	93
Figure 15: Scatterplot matrix of four quality-related measures for <i>lab_hand_FOF</i>	95
Figure 16: Scatterplot matrix of four experimenter-selected measures (‘Hand-picked C’) for <i>lab_htk_FOF</i>	95
Figure 17: Scatterplot matrix of four experimenter-selected measures (‘Hand-picked C’) for <i>lab_hand_FOF</i>	96
Figure 18: Smoothed f0 contour and waveform from an elicited utterance of “than I did” (subject focus) in an interrogative context.....	98
Figure 19: Smoothed f0 contour and waveform from a naturally occurring	

utterance of “than I did” (subject focus) in a declarative context.....	98
Figure 20: Distributions of listener accuracies (web data).....	110
Figure 21: Distributions of listener balanced error rates (web data).....	111
Figure 22: Distribution of items from <i>web2</i> according to listener accuracy rate.....	112
Figure 23: Distribution of listeners according to listener accuracy rate. First occurrence focus.....	117
Figure 24: Distribution of listeners according to listener balanced error rate. First occurrence focus.....	118
Figure 25: Distribution of listeners according to listener accuracy rate. Second occurrence focus.....	119
Figure 26: Distribution of listeners according to the listener balanced error rate. Second occurrence focus.....	119
Figure 27: Distribution of items from laboratory production experiment according to listener accuracy rate. First occurrence focus.....	120
Figure 28: Distribution of items from laboratory production experiment according to listener accuracy rate Second occurrence focus.....	121
Figure 29: Distribution of speakers from laboratory production experiment according to listener accuracy rate.....	122

Chapter 4

Figure 1: Pitch tracks and MAE_ToBI annotations from Ahn (2008).....	168
--	-----

Chapter 5

Figure 1: Smoothed f_0 track (in Hz) of (20). Clausal focus configuration.....	212
Figure 2. Smoothed f_0 track (in Hz) of (21). Clausal focus configuration.....	214
Figure 3. Smoothed f_0 track (in Hz) of (22). Clausal focus configuration.....	215

Figure 4. Smoothed f0 track (in Hz) of (27). Clausal focus configuration.....	218
Figure 5. Smoothed f0 track (in Hz) of (28). Clausal focus configuration.....	220
Figure 6. Smoothed f0 track (in Hz) of (31). Argument focus configuration.....	223
Figure 7. Smoothed f0 track (in Hz) of (32). Argument focus configuration.....	225
Figure 8. Smoothed f0 track (in Hz) of (41). Adnominal ER focus configuration.....	230
Figure 9. Smoothed f0 track (in Hz) of (42). Adnominal ER focus configuration.....	232
Figure 10. Smoothed f0 track (Hz) of (43). Adnominal ER focus configuration.....	233
Figure 11. Smoothed f0 track (in Hz) of (51). Sub-clausal focus configuration.....	237
Figure 12. Smoothed f0 track (in Hz) of (52). Sub-clausal focus configuration.....	239
Figure 13. Smoothed f0 track (in Hz) of (53). Sub-clausal focus configuration.....	240
Figure 14. Smoothed f0 track (in Hz) of (52). Sub-clausal focus configuration.....	241
Figure 15. Average f0 topline for four different focus conditions (Eady et al. 1986:245).....	251
Figure 16. Smoothed f0 track of (52). Double focus configuration (adnominal ER, predicate).....	255
Figure 17. Smoothed f0 track (in Hz) of (52). Double focus configuration	

(adnominal ER, predicate).....	257
Figure 18. Smoothed f0 track of (74). Double focus configuration (adnominal ER, predicate).....	259
Figure 19. Smoothed f0 track (in Hz) of (75). Double focus configuration (subject, predicate).....	261
Figure 20. Smoothed f0 track (in Hz) of (76). Double focus configuration (subject, predicate).....	262
Figure 21. Smoothed f0 track of (88). Predicate focus configuration.....	269

LIST OF TABLES

Chapter 1

Table 1. Comparisons of the agreement rate of prosody annotation following the ToBI transcription conventions reported in different studies (Mo 2010:8).....	13
--	----

Chapter 2

Table 1. Files from a retrieval with target “in my opinion”.....	46
Table 2. The most frequent domain names in the <i>in-my-opinion</i> run.....	47

Chapter 3

Table 1. Summary of datasets analyzed.....	76
Table 2. Summary of training set / test set pairs.....	76
Table 3. List of feature sets selected for training on dataset <i>web1</i>	78
Table 4. Accuracy rate (and balanced error rate) for different models: training set <i>web1</i> ; test set <i>web2</i>	84
Table 5. Accuracy rate (and balanced error rate) for different models: training set <i>web1</i> ; test set <i>lab_htk_FOF</i>	85
Table 6. Accuracy rate (and balanced error rate) for different models: training set <i>web1</i> ; test set <i>lab_hand_FOF</i>	86
Table 7. Accuracy rate (and balanced error rate) for different models: training set <i>web1</i> ; test set <i>lab_htk_declarative</i>	99
Table 8. Accuracy rate (and balanced error rate) for different models: training set <i>web1</i> ; test set <i>thanIdidlab_interrogative</i>	100
Table 9. Accuracy rate (and balanced error rate) for different models: training set <i>web1</i> ; test set <i>lab_htk_SOF</i>	101

Table 10. List of feature sets selected for training on dataset	
<i>lab_htk_FOF</i>	104
Table 11. Accuracy rate (and balanced error rate) for different models:	
training set <i>web1</i> ; test set <i>lab_htk_FOF</i>	105
Table 12. Accuracy rate (and balanced error rate) for different models:	
training set <i>web1</i> ; test set <i>lab_htk_FOF</i>	106
Table 13. Summary of generalized linear mixed models for listener responses to a subset of <i>web2</i> using predictors from hand-selected feature sets <i>Hand-picked A</i> and <i>Hand-picked C</i>	113
Table 14. Summary of Analysis of Variance (ANOVA) between the two mixed- effect models in table 13.....	114
Table 15. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models (fixed effects from featureset <i>Hand-picked A</i>) with and without item as random effect.....	114
Table 16. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models (fixed effects from featureset <i>Hand-picked A</i>) with and without participant as random effect.....	115
Table 17. Summary of generalized linear mixed models for listener responses to a subset of <i>lab_SOF</i> using predictors from hand-selected feature sets <i>Hand-picked A</i> and <i>Hand-picked C</i>	123
Table 18. Summary of generalized linear mixed models for listener responses to a subset of <i>lab_SOF</i> using predictors from hand-selected feature sets <i>Hand-picked A</i> and <i>Hand-picked C</i>	124
Table 19. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models for FOF data (cf. Table 17).....	125

Table 20. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models for SOF data (cf. Table 18).....	125
Table 21. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models (fixed effects from feature set <i>Hand-picked A</i>) for FOF data with and without participant as random effect.....	126
Table 22. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models (fixed effects from feature set <i>Hand-picked A</i>) for FOF data with and without item as random effect.....	126
Table 23. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models (fixed effects from feature set <i>Hand-picked A</i>) for SOF data with and without participant as random effect.....	126
Table 24. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models (fixed effects from feature set <i>Hand-picked A</i>) for SOF data with and without item as random effect.....	126
Table 25. Summary of categorical prominence contrasts and related examples.....	136

Chapter 4

Table 1. Partial classification of focus sensitive expressions modified from Beaver & Clark (2008:78)	159
--	-----

CHAPTER 1

INTRODUCTION

1 Overview of the dissertation

In this dissertation, I present a new approach to research on meaning and prosody, using speech “harvested” from the web. To my knowledge, this is the first linguistic study using speech found online. With the Internet more than twenty years old¹, one might justifiably be surprised. Certainly, linguists have been using the web as a written corpus for almost as long as there have been search engines².

Steady advances in speech to text technologies have made the study of web speech possible—in particular automatic speech recognition (ASR)—and the work in this dissertation is indebted to the many people who have worked at developing them. Ramp, formerly Everyzing, is one of the first companies to apply these technologies to large quantities of web data. Mats Rooth and I have developed a set of command-line tools to interact with Ramp-powered search interfaces, to download speech data from content providers and to process the speech data for linguistic research. I describe this method of harvesting and quantify its efficacy in Chapter 2.

Speech data from the web are distinct from other available kinds of speech data in many ways. The data I harvest from the web are naturally occurring—mostly spontaneous and unscripted; they are diverse—uttered by speakers of different ages, from different socio-economic backgrounds and geographical regions and in different real-world contexts; and they are partly controlled for syntactic and phonological context—each token in a dataset is an utterance of the same string of text. This is a unique combination of properties which I believe

¹ Tim Berners-Lee’s (1989) proposal “Information Management: A Proposal” is largely credited as the blueprint for the world wide web. Berners-Lee and Robert Cailliau implemented the first internet communication between a server and a browser on December 25, 1990.

² WebCrawler (1994) was the first world wide web search engine that allowed full-text search. Mihalcea & Moldovan (1999) and Resnik (1999) were the first researchers to present results from web research at the Association for Computational Linguistics (Kilgariff & Grefenstette 2003).

opens up many new lines of inquiry. The linguistic studies in this dissertation, particularly Chapters 3 and 5, would not have been possible using other sources of data alone.

The approach I am advocating is not to abandon previous methods of data acquisition, however. Like all types of data, speech harvested from the web has both advantages and disadvantages. For example, the recordings are of variable recording quality, often contain disfluencies and may be biased towards more clearly articulated speech because they are indexed using ASR. I take it as uncontroversial that for most phenomena, a thorough investigation requires more than one methodology, often applied iteratively. In Sections 3 and 4 of this introduction, I advocate a pluralistic view of linguistic data and methodology and argue that web-harvested speech data fill an important gap in our methodological toolbox.

English speakers use prosody (e.g. intonation, stress, rhythm) to convey many facets of linguistic meaning, including affect, social identity, cognitive processing and syntactic structure. I hope that the particular harvest methodology and the general empirical approach will be useful for the study of all of these. In this dissertation, however, I am interested in one particular meaning-prosody correlation, between

- (1) (i) which information in a discourse is most important; and
 (ii) which material in an utterance is realized with prosodic prominence

which I will mostly refer to as “focus”.

The study of this correlation comes with many challenges, both empirical and methodological. The first and perhaps most fundamental has been to determine the objects of study, over which there remains considerable debate. And a major methodological challenge has been bridging the several subdisciplines and research cultures that have approached different aspects of this correlation from different perspectives. Linguists of different orientations—phoneticians, phonologists, syntacticians, semanticists, pragmaticists—as well as philosophers, psychologists and computer scientists all have a vested interest in understanding this correlation, yet collaboration and communication among them is the exception, rather than the rule.

In Section 2, I describe the correlation (1), and its empirical and methodological challenges in more detail, and with a broad audience in mind. In the course of the exposition, I also introduce certain theoretical frameworks presupposed in later chapters, including the autosegmental-metrical view of prosody, the alternative semantics for focus and a theory of focus anaphoricity.

The remaining chapters of the dissertation are organized as follows. In Chapter 2, I describe the method of harvesting speech data from the web (cf. Howell & Rooth 2009), quantify its efficacy and discuss possible improvements.

Chapter 3 concerns the location and prosodic realization of focus in comparative clauses (e.g. *than I did*). Using machine learning and human classifiers, I discover a robust correlation between sets of acoustic cues of prosodic prominence and the location of focus predicted by linguistic theory. Based on the robustness of non-intonational acoustic cues, I hypothesize that focus may be realized by discrete, paradigmatic (i.e. cross-utterance) categories of stress.

The correlation between acoustic cues and focus location obtains in both the web-harvested corpus data and the laboratory-elicited experimental data. I take advantage of the experimental paradigm to study two conditions not available in the web corpus data: so-called “second occurrence” focus (e.g. repeated focus) and focus in interrogative contexts. The results confirm that intonational measures are not significant predictors of focus location in either of these conditions. Non-intonational cues are significant for both conditions, but slightly weaker and more variable in the second occurrence focus condition compared to regular first occurrence focus. The results are consistent either with optional rather than obligatory realization of second occurrence focus or a failure of the experimental paradigm to elicit genuine second occurrence focus consistently.

In Chapter 4, I introduce the adnominal emphatic reflexive (adnominal ER; e.g. *himself* in *Jane met Chomsky himself*). I argue that it is an instance of a theoretically possible but largely unattested (until now) focus-sensitive operator with sub-propositional scope. Using constructed

data and personal introspection, I argue that the adnominal emphatic reflexive exhibits most of the expected pragmatic, semantic, syntactic and prosodic properties of focus sensitive operators and that those purported properties it does not hold turn out not to hold of all focus sensitive operators.

Analyzing the adnominal ER as a sub-propositional operator frees the meaning of the adnominal ER from focus structure at the level of the sentence. The adnominal ER may but need not have sentential focus. I am thus able to reconcile opposing approaches to the semantics of this construction: one which maintains that the adnominal ER is a focus-sensitive operator and another which maintains that it is itself focused.

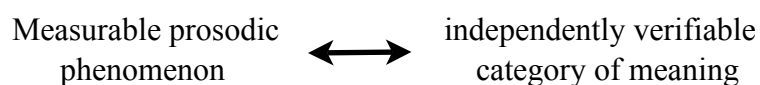
In Chapter 5, I explore the predictions of this “extended” focus-sensitive operator analysis. I analyze the discourse context and intonation of individual web-harvested data: utterances containing the string *he himself*. I identify naturally-occurring examples of 7 different focus configurations with distinct discourse licensing conditions and intonational realizations. This finding is evidence against the hypothesis of focal determinism, according to which certain linguistic constructions in a language are inherently or obligatorily focused.

2 *Objects of study*

2.1 *Correlates of prosodic prominence*

The simplest account of the correlation (1) posits a direct mapping between a single, measurable prosodic correlate and a single, independently verifiable category of meaning. Theoretically, this is a kind of null hypothesis. And for practical concerns, such as the automatic detection of meaning or the synthesis of prosody, it would seem reasonable to push the idea as far as possible.

(2) Direct mapping



Beginning with prosodic prominence, we find that several phonetic correlates have been proposed. Acoustically, for example, prominence may be realized with some combination of fundamental frequency (f_0), relative intensity, vocal tract resonances (formants and spectra) and duration (e.g. of acoustic events, of speech segments, of syllable nuclei, of words). Perceptually, these correspond roughly to pitch, loudness, vowel quality and length. Articulatorily, prominence has been claimed to be realized by increased respiratory effort (e.g. more air expelled from the lungs), changes in the glottis (the muscles known as “vocal folds” and the space between them), exaggerated supralaryngeal articulation (changes in the vocal tract above the glottis, including the tongue, teeth and lips) and increased jaw displacement.

Acoustic measures of prominence are in many respects the easiest to obtain. Methods of signal processing, while still imperfect, are well established and can be applied on any personal computer using free software such as Praat (Boersma 2001). For example, an “auto-correlation” method detects regularities in a sound wave in order to calculate f_0 over time (aka a “pitch track”). This method works best for vowels and sonorant consonants (e.g. [l] and [m]) which have periodic waves; the method has difficulty with stop consonants (e.g. [p] and [t]) and fricatives (e.g. [f] and [s]) which have no or aperiodic waves. Figure 1 displays the waveform (top) and pitch track (bottom) for an utterance of “I implied that you”. The f_0 contour is disrupted by the silence of the consonants [p] and [t] and the aperiodicity of the fricative /θ/ and of aspiration in the release of [t]. The speaker also laughs while speaking the word *implied*, which results in aperiodicity on the vowel [aɪ]; this is observable as a change in both the waveform and the pitch track.

Articulatory measures require more invasive and often costly methods, such as palatography, ultrasound, laryngography or aerometry, and are necessarily available only for controlled, laboratory speech. Some auditory measures, such as auditory frequency, are related to acoustic measures directly; typically, however, perceptual studies measure prosodic correlates

indirectly by observing listener behavior, such as discrimination or identity rating, response speed or eye gaze. While these are useful methods, additional human and/or instrumental resources are required compared with acoustic analysis and one must also take special care to tease apart the effects of prosodic prominence from other linguistic and non-linguistic factors.

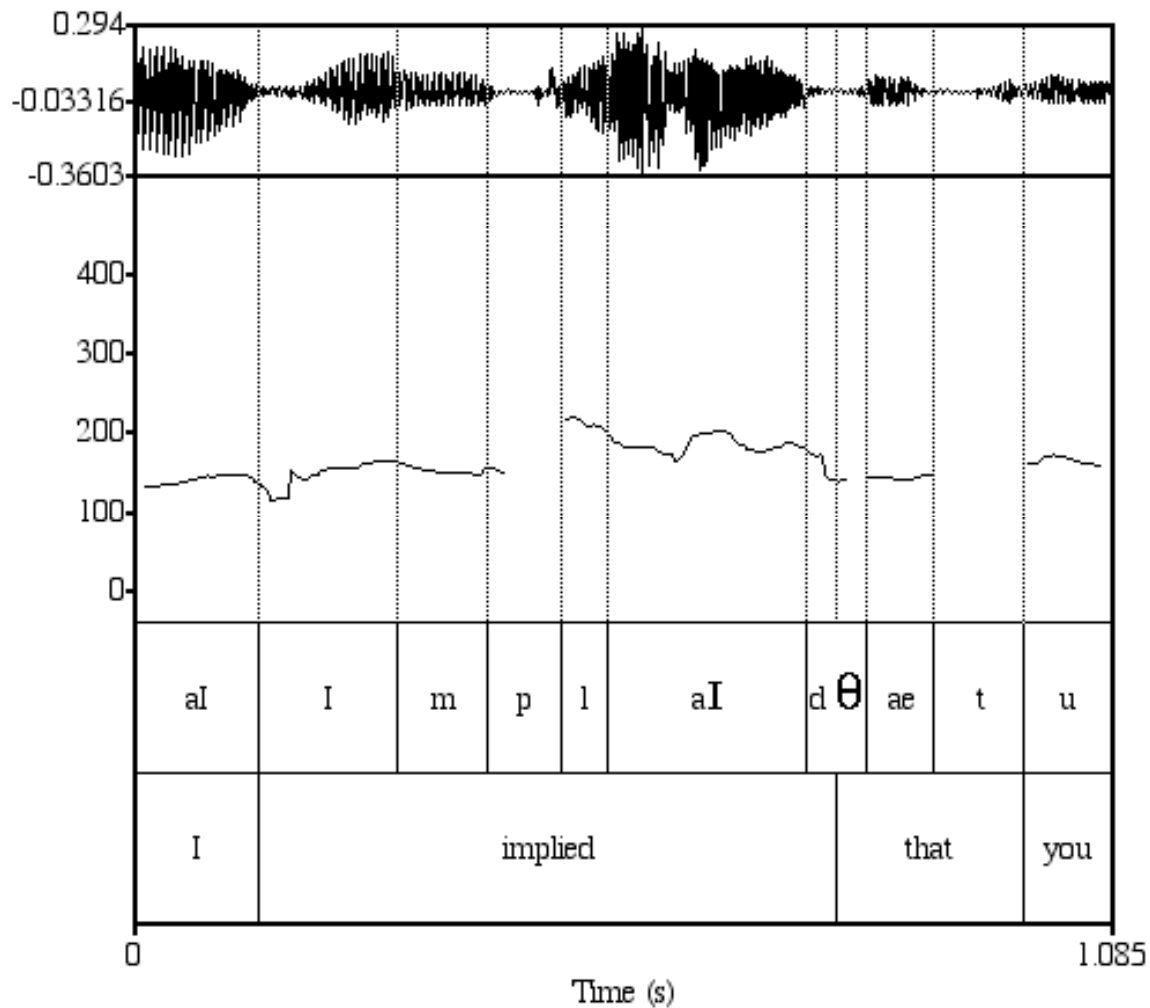


Figure 1. Example waveform (top; measured in dB) and f0 contour (bottom; measured in Hz) for an utterance of “I implied that you”.

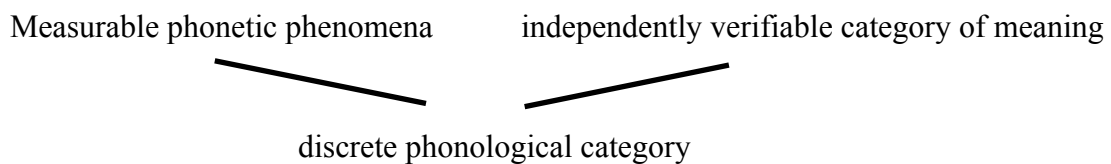
Some scholars have pursued the direct mapping hypothesis (2) by seeking specific acoustic correlates of semantic-pragmatic categories like focus (Cooper *et al.* 1985; Eady *et al.* 1986) or

givenness (Fowler & Housum 1987; Shields & Balota 1991; Koopmans-van Beinum & van Bergem 1989). They are able to find some robust statistical correlations between measurements of the acoustic signal and categories of meaning.

Yet as Ladd (1996) emphasizes, it does not follow from such results that the correlation between meaning and prosody lacks grammatical mediation. Indeed, grammatical mediation obtains in other domains of language, as witness the impressive body of research in segmental phonology. Moreover, the existence of discrete prosodic categories (as distinct from the continuous acoustic signal or kinematics of speech articulators) is an empirical question which ought to be explored, or at least controlled for, rather than ignored.

If prosodic prominence is phonologically mediated, then the mapping would be indirect, as in (2). In this chapter, I will consider two phonological categories as possible correlates of prosodic prominence: stress and pitch accent.

(2) Phonologically mediated mapping



While most linguists and laypeople have an intuition that there is a category stress, it has been at the center of theoretical debates for more than 250 years (Ladd 1991) and even after decades of instrumental investigation, its grounding in phonetics remains uncertain. As Hayes (1995) proclaims, the definition of stress has been “one of the perennially debated and unsolved problems of phonetics”.

A large part of the problem is that stress appears to subsume a disparate set of phonetic phenomena. All of the acoustic and articulatory parameters discussed so far have been implicated in the realization of stress and a considerable body of work in instrumental phonetics has failed to uncover a single unique invariant set of phonetic correlates. The research of Mo (2011)

suggests that in fact each individual speaker may use a different set of phonetic cues. Conversely, any of these individual phonetic parameters may be a channel for multiple percepts. For instance, duration—in addition to signaling stress—is a cue to vowel length, the voicing of fricatives, prosodic phrasing and the voicing of postvocalic consonants and is influenced by extragrammatical properties such as speaker identity and rate of speech (Klatt 1976, Turk *et al.* 2006).

Intuitions of stress have been equally problematic. Native speakers produce stress effortlessly and can detect incorrect stress produced by non-native speakers. Yet, as with other grammatical phenomena, speakers typically lack direct access to this implicit knowledge. Hayes (1996) speculates that prosodic phenomena may in fact be the “least accessible to consciousness”.

To compensate for the difficulty of the analytic task, one might rely on linguistically trained experts to identify stress, or average across the responses of several naïve listeners. Mo (2010) follows the latter strategy, identifying prominence on individual words in an utterance by asking listeners to annotate the words as either prominent or non-prominent, and then calculating a probabilistic prominence score for each word based on the proportion of listeners who annotated it as prominent. Mo’s transcription scheme presupposes that stress (or some more general notion of prominence) is binary—present or absent—for the purpose of elicitation, but produces a representation which is continuous.

As for the theory of stress, analysts have viewed it alternately as a local property of syllables or words, on the one hand, and as a structural relation between items in the same utterance, on the other: the paradigmatic and syntagmatic views, respectively.³ On the paradigmatic view—which Ladd (1991) describes as the “commonsense” view—holds that a syllable may either be stressed or unstressed, and that there may be more than one level of stress. One finds this view represented in phonetic alphabets, such as the International Phonetic

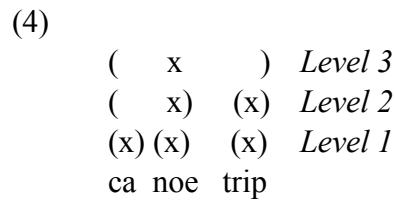
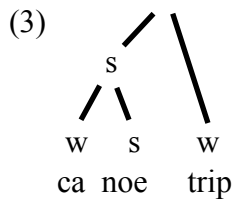
³ Fox (2000) attributes the source of this distinction to Saussure 1916.

Alphabet (IPA); in the work of many theoretical linguistics working on syntax and semantics who use typographical strategies like capitalization or italicization to indicate stress; and adopted as word or syllable features in early generative theories and in computational approaches. On this view, stress is interpreted in absolute terms—identified by inspection or compared against alternate realizations.

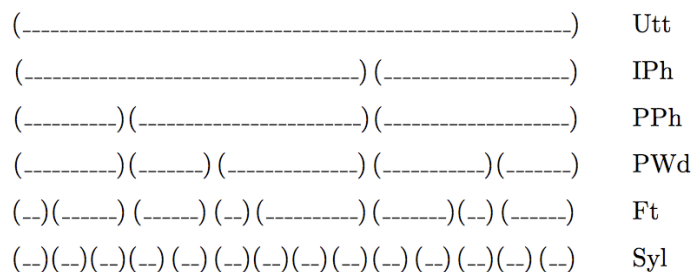
On the other hand, linguists have long noted that stress, and prosodic prominence generally, is also relational, identified by comparison of items within the same utterance (e.g. Saussure 1967[1916]; Jakobson, Fant, Halle 1951; Lehiste 1970; Ladefoged 1975). This observation fit awkwardly into early phonological frameworks which were concerned primarily with the paradigmatic comparison of segmental phenomena, such as allophonic variation. In early literature in generative phonology, for example, one finds representations of stress as a feature (e.g. [\pm heavy], [\pm stress]) or stress degrees (e.g. {1,2,3,4...}).

Metrical phonology (e.g. Liberman 1975, Liberman & Prince 1977; Hayes 1981; Giegerich 1983; Prince 1983; Selkirk 1984; Halle & Vergnaud 1990) views stress as hierarchically organized rhythmic structure. Linguists working in metrical phonology developed new, non-linear representational models, including the tree notation (e.g. 3) and the grid notation (e.g. 4). The tree notation captures stress as a binary (and on some accounts n-ary; e.g. Beckman 1984) relationship of relative rhythmic strength: s(trong) and w(eak). One disadvantage of this notation is that one can no longer express paradigmatic stress levels.⁴ For example, we have the intuition that the first syllable of *canoe* is weaker than *trip* in (3-4): only the former may undergo vowel reduction. The grid notation was introduced to remedy this and other problems: stress levels are represented as column height; prosodic constituency may be represented with bracketing (Halle & Vergnaud 1987; Idsardi 1992).

⁴ Liberman & Prince re-introduce a feature [\pm stress] to capture such cases.



Selkirk (1978, 1980) introduced labels for each level of prominence, relating them to traditional notions such as the metrical foot. According to prosodic hierarchy theory, each phonological domain has distinct and intrinsic phonological and phonetic properties, although one finds differing views on the number of prosodic categories and their properties.



(5) Prosodic hierarchy (Selkirk 1986:384)

For example, one may identify stress at the level of the prosodic word by observing other phonological phenomena, such as vowel reduction and onset deletion (e.g. Selkirk 1978,1996, 2008a).

- (6)
- | | | | |
|----|-------------------------|-------|-----------------|
| a. | <i>at</i> (stressed) | [æt] | vowel reduction |
| | <i>at</i> (unstressed) | [ət] | |
| b. | <i>him</i> (stressed) | [hɪm] | onset deletion |
| | <i>him</i> (unstressed) | [m] | |

Experimentally, the evidence from metrical phonology suggests that one ought to observe the production and perception of stress both paradigmatically and syntagmatically, although it is

far from obvious how one ought to normalize a given measurement and the method will be determined in part by the data.

The evidence from metrical phonology also goes some way to explaining why no invariant set of phonetic cues has been found to correlate reliably with discourse prominence: prominence is grammatically mediated and inherently relational.

A second important candidate for phonological correlate of prominence is pitch accent. Many different grammatical representations of intonation have been proposed. For example, British schools of intonation (e.g. Crystal 1969; Cruttenden 1986) have viewed intonation in terms of tonal contours (e.g. falling-rising); while American structuralists (e.g. Pike 1945; Wells 1945; Trager & Smith 1951) viewed intonation as a four-level system of pitch phonemes. The notion of a pitch accent was originally proposed for English by Bolinger (1958a). Ladd (1996) offers the following characterization: “a local feature of a pitch contour—usually but not invariably a *pitch change* and involving a local minimum or maximum—which signals that the syllable is *prominent* in the utterance”.

Pitch accent has come to be most closely associated with the work of Janet Pierrehumbert and her colleagues and the autosegmental-metrical theory of prosody which developed from her 1980 dissertation, although this theory has a number of sources, including Bruce’s (1977) analysis of Swedish tones, and autosegmental models of English intonation in Goldsmith (1976), Liberman (1975) and Leben (1975).

Pierrehumbert’s (1980) dissertation was further developed in Beckman & Pierrehumbert (1986) and was the basis of the annotation of intonation in the Tones and Break Indices (ToBI) annotation standard (Silverman *et al.* 1992, Beckman & Ayers 1994). In these systems, English has two basic tones—high (H) and low (L). These tones may align with stressed syllables (in which case they are “pitch accents”) or they may align to the edges of prosodic boundaries (in which case they are boundary tones). Pitch accents may be simple (H* and L*) or bitonal (H*+L, H+L*, L*+H, L+H*); the stress-aligned tone is notated with the asterisk (*) and the

secondary tone is represented by a preceding or following plus (+). A boundary tone (H% or L%) signals the edge of an intonational phrase; a phrase accent (H or L) signals the edge of an intermediate or phonological phrase (cf. Beckman & Pierrehumbert 1986; Hayes & Lahiri 1991).

Two canonical tone sequences are illustrated in (7-8), although the intonational grammar allows many others. In a typical declarative utterance, stressed syllables have an H* pitch accent and a L% boundary tone (cf. 7). In a typical interrogative utterance, stressed syllables have a L* pitch accent and a H% boundary tone (cf. 8).

- (7) Legumes are a good source of vitamins.
(cf. Legumes are a bad source of vitamins.)
- (8) Legumes are a good source of vitamins?
(i.e. Are you kidding? They're an awful source of vitamins.)

Unlike stress, the category of pitch accent has an obvious phonetic channel, namely *f0*. Nonetheless, a straightforward phonetic correlate of pitch accent has been equally elusive. The *f0* contour (cf. pitch track) produced by autocorrelation algorithms of signal processing software may at best be described as a “very rough first-pass phonetic representation” (Beckman & Venditti 2010). As discussed above, the algorithm fails in the absence of periodic voicing; the *f0* contour is also sensitive to so-called micro-variation: small changes due to the influence of different segments (e.g. Peterson 1986; Silverman 1986). Simply measuring *f0* extrema is therefore not reliable.

As for the perception of pitch accent, one finds that expert transcribers trained on the ToBI system show considerable inter-annotator agreement for pitch accent (cf. Table 1), although these annotators have access to a visual representation of the speech signal (i.e. waveform, spectrogram and f_0 contour) and may listen to an utterance multiple times before reaching a decision. Clearly, the context for annotation differs from the natural parsing of intonation by

naïve listeners; it is unclear whether the annotators' performance reflects an accurate representation of listener perception. Expert annotation of elicited or corpus data is also labor-intensive and time-consuming.

Agreement Rate	Pitrelli <i>et al.</i> (1994)	Syrdal & McGory (2000)	Yoon <i>et al.</i> (2004)	Dilley <i>et al.</i> (2006)
Prominence labeling	81%	91%	87%	87%

Table 1. Comparisons of the agreement rate of prosody annotation following the ToBI transcription conventions reported in different studies (Mo 2010:8)

As a potential correlate of discourse importance, the notion of pitch accent alone is also inadequate. There is no one-to-one correspondence between pitch accent and any category of meaning, such as contrast or novelty. In (9b), John is introduced into the discourse by the question *Who did John's mother praise?*. Yet, the pronoun *him*, which refers to John, is realized in the answer with a pitch accent. And uttered as all new information, (10) does not signal any contrast between Manny and other individuals or between the annex and other locations.

Similarly, there is no one-to-one correspondence between pitch accent and, for example, givenness. The words *hired* and *work* in (10) may optionally be realized without a pitch accent (particularly in fast speech) without signaling that they are in any sense old in the discourse.

- (9) a. Who did John's mother praise? (from Schwarzschild 1999)

H* L-L%

- b. She praised him.

- (H*) H* (H*) H*
(10) We hired Manny to work on the annex. (from Selkirk 2008a)

In another move, we may appeal to a pitch accent having a special status. Most traditions of research on intonation recognize a most prominent and/or obligatory part of the intonational contour: in the British tradition (e.g. Palmer 1922) it is called the "nucleus"; in Halliday (1967),

the “tonic”; in Pike (1945), the “primary contour”; in Hockett (1958), the “head”. In the autosegmental-metrical systems, the nuclear accent has no formal status but may be identified as the last pitch accent in a phonological phrase.

Any relationship between nuclear pitch accent and discourse importance is, however, indirect. There is a considerable literature on a phenomenon known as focus projection or focus ambiguity (see e.g. Selkirk 1984; Cinque 1993; Zubizarreta 1998; Legate 2002). A nuclear pitch accent (or the related concept of nuclear stress, cf. Newman 1946; Chomsky & Halle 1968) may signal the discourse importance of the individual word on which it is realized; it may also signal the discourse importance of a larger syntactic constituent to which it belongs to. For example, Selkirk (1984) argues that the utterance in (11) with a nuclear accent on *bats*⁵ is an appropriate response to any of the questions in (12).

- (11) H* (H*) H* L-L%
 Mary bought a book about bats
- (12) a. What did Mary buy a book about?
 b. What kind of book did Mary buy?
 c. What did Mary buy?
 d. What did Mary do?
 e. What’s been happening?

A related difficulty in mapping nuclear accent and discourse importance is that, although a nuclear pitch accent is perceived as more prominent relative to other pitch accents (*viz.* more prominent than those in the same phonological phrase), it does not follow that the nuclear pitch accent will be realized, in the case of high toned pitch accents, with a higher *f0*. In a phenomenon known as downstep or catathesis, a series of high tones within a particular domain, such as the phonological phrase are realized in a “terraced contour” (O’Connor & Arnold 1973) or “staircases” (Lieberman & Pierrehumbert 1984), illustrated in Figure 2.

⁵ The pitch accent annotation is mine.

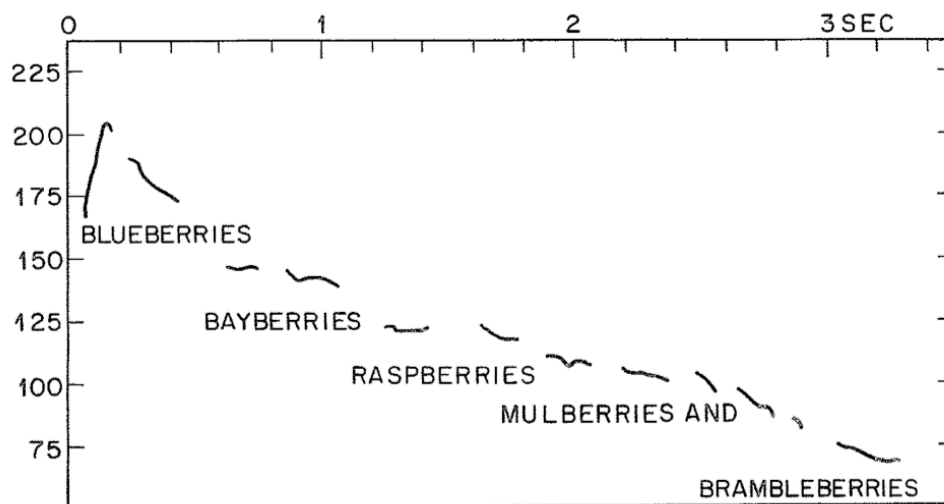


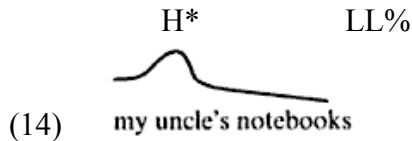
Figure 2. An f_0 contour with a series of downstepped H^* pitch accents from Liberman & Pierrehumbert (1984)

In Chapter 5, I use this phenomenon as a criterion for distinguishing different domains of focus, according to the following two properties: (i) downstep applies within the phonological domain corresponding to a focused constituent; and (ii) a focused constituent causes the f_0 topline to be “reset”. Consider the following minimal pair from Ladd (1996).⁶ (13a) is an appropriate response to “Which of your uncle’s possessions did you find in the attic?”; (13b) is an appropriate response to “What did you find in the attic?”. Phonologically, (13ab) differ only in whether the pitch accent on *notebooks* is downstepped (noted with the symbol !).

- (13) a. $\overset{H^*}{\text{my}} \overset{H^*}{\text{uncle's}} \overset{LL\%}{\text{notebooks}}$ b. $\overset{H^*}{\text{my}} \overset{!H^*}{\text{uncle's}} \overset{LL\%}{\text{notebooks}}$

A rendition of this utterance without any pitch accent on *notebooks*, cf. 14, receives yet another interpretation: e.g. a response to the question “Whose notebooks did you find in the attic?”.

⁶ The pitch accent annotation and contexts are mine.



Although the phenomenon of downstep is generally acknowledged, there has been no consensus on whether downstepped pitch accents are pragmatically conditioned (Ladd 1993) or phonologically conditioned (e.g. Pierrehumbert 1980, Beckman & Pierrehumbert 1986) and whether downstepped accents are in fact perceived as categorically distinct (cf. Dainora 2001; Yoon 2007).

2.2 *Correlates of discourse importance*

In this introduction, I have used the intentionally vague term “discourse importance” to cover the particular notion of focus that I assume in the body of the dissertation, as well as the many related concepts and phenomena found in the linguistic and psychological literature. Scholars of speech and of linguistic meaning both face the usual difficulties of negotiating different research traditions and schools of analysis; scholars of linguistic meaning are additionally challenged by the lack of a directly observable and quantitatively measurable physical object, such as the acoustic signal.

In attempting to identify importance in a given utterance, one may for example observe the content of prior discourse. We may search previous utterances for an explicit contrast or a previous mention which would license part of the utterance to be marked as (not) important. In (15), for example, *Daddy* in the second sentence contrasts in with *Grandma* in the first and *often* contrasts with *never*; these words are realized with prosodic prominence (indicated typographically with capitalization). The predicate *drives 200 m.p.h.* is introduced in the first sentence and therefore is old or given information in the second; and we intuit that it is realized without prosodic prominence. Similarly, in (16), *John* contrasts with *my brother* and *Milan* contrasts with *Naples*, and we intuit that both are realized with prosodic prominence. The string

traveled to is introduced in the first sentence and is therefore old or given in second; we intuit that it is realized without prosodic prominence.

- (15) Grandma often drives 200 m.p.h.; DADDY NEVER drives 200 m.p.h.
- (16) My brother traveled to Milan; JOHN traveled to NAPLES.

Consider now the minimally different discourses in (17-18). In (17), the predicate *goes over the speed limit* is introduced into the discourse for the first time. Yet, we somehow interpret it as old or given in the same way that we interpret *drives 200 m.p.h.* in (15); and we intuit that it is realized without prosodic prominence. This interpretation is possible because we know that 200 m.p.h. is above the speed limit, and the speaker (or writer) assumes that we know this or that we can infer it. Similarly, in (18), *south* is contrastive even in the absence of an explicit contrast (e.g. north), and we intuit that it is realized with prosodic prominence; and *Italy* is interpreted as old or given even though it is uttered for the first time, and we intuit that it is uttered without prosodic prominence. This interpretation is possible because we know that Milan is in the north of Italy and the speaker (or writer) assumes that we know this or that we can infer it. This phenomenon is known as bridging (cf. Clark & Haviland 1974; Clark & Haviland 1977).

- (17) Grandma often drives 200 m.p.h.; DADDY NEVER **goes over the speed limit**.⁷
- (18) My brother traveled to Milan; JOHN traveled to the **SOUTH of Italy**.

Recognizing a contrast between *my brother* and *John* in (16,18) also depends on the identity of the speaker. Uttered by John's brother, *John* does not contrast and we intuit that it must be realized without prosodic prominence as in (19). This sensitivity to speaker identity is an instance of indexicality.

- (19) My brother traveled to Milan; John traveled to the SOUTH of Italy, too.
[uttered by John's brother]

⁷ Example from Deemter 1999.

Phenomena like bridging and indexicality preclude a simple analysis of discourse importance calculated directly from previous utterances. Rather, a successful analysis must account for the role of real world context (e.g. location, time, interlocutor identity) and the assumptions that speaker and other interlocutors make about their beliefs and intentions.

Add to this empirical situation that the literature on discourse importance, as such, is characterized by a lack of consensus on the notion of discourse importance. Here I will discuss three broad distinctions that inform many accounts of discourse importance:

- (20) (i) grammatical and extra-grammatical;
(ii) pragmatic and semantic;
(iii) novelty/givenness and contrast.

Grammatical phenomena are governed by sets of rules, constraints or conventions that comprise our linguistic knowledge; extra-grammatical phenomena are governed by principles which are not specific to language, although they may affect language.

For example, the psychological notion of focus of attention (e.g. Bosch 1988; Dahl & Gundel 1981; Garrod & Sanford 1982; Grosz & Sidner 1986; Gundel, Hedberg & Zacharski 1993)⁸ is concerned with general cognitive properties and processes of memory and attention state of speaker and hearer in relation to an entity. Psychological focus may influence several linguistic phenomena, such the appropriate use of pronoun forms, as in (21-22) (Gundel 1999). The bull mastiff is introduced or “activated” in both (21) and (22), but only in (21) is it the psychologically focused.

- (21) a. My neighbor’s bull mastiff bit a girl on a bike.
b. It’s / That’s the same dog that bit Mary Ben last summer.
- (22) a. Sears delivered new siding to my neighbors with the bull mastiff.
b. #It’s / That’s the same dog that bit Mary Ben last summer.

⁸ Cited in Gundel (1999)

I follow Kadmon (2001) in separating the distinction (i) between what's in the grammar and what's outside of the grammar from the distinction (ii) between semantics and pragmatics. Let's say that semantics is concerned with "literal meaning": "what is said" (Grice 1975) or what is "at issue" (Potts 2005). In formal semantic theory, the rule system of semantics compositionally interprets syntactic expressions (e.g. phrases and sentences). Pragmatics is concerned with use of language "beyond literal meaning" (Kadmon 2001): "what is meant" (Grice 1975) and how the information is structured in the service of conversation goals. The rule system of pragmatics interprets semantic expressions (e.g. propositions) in a particular context of use (i.e. an utterance).

Evidence for the grammatical mediation of discourse importance comes from the sensitivity to discourse importance shown by certain linguistic expressions. Dretske (1972) offers the following argument:

... contrastive differences ... however one may choose to classify them, are significantly involved in determining the meaning (hence, semantics) of a variety of larger expressions in which they are embedded. If $C(U)$ is a linguistic expression in which U is embedded, and U can be given different contrastive foci (say U_1 and U_2), then it often makes a difference to the meaning of $C(U_1)$ and $C(U_2)$, will have to be provided with resources for distinguishing between U_1 and U_2 . (Dretske 1972)⁹

For example, the sentence in (23) is ambiguous between at least two interpretations. In formal semantics, meaning is understood according to truth conditions: what the world would have to be like in order for a sentence to be true (independently of what the world is in fact like). We can imagine a world in which Cristina is an anthropologist and writes about many cultural topics including cooking; she herself is in fact a terrible cook and consciously avoids the kitchen. Uttered in this world, we would judge (23a) as true, and (23b) as false. Consider another world in

⁹ Cited in Rooth (1996b)

which Cristina is a celebrity chef and has published many books of recipes and cooking technique; she is an avid reader of science fiction novels, but has never attempted to write one. In this world, we would judge (23a) as false and (23b) as true. Truth-conditionally, then, (23a) and (23b) do not have equivalent meanings. In this sense, discourse importance is semantic.

- (23) Cristina **only** writes about cooking.
 - a. Cristina **only** WRITES about cooking.
 - b. Cristina **only** writes about COOKING.

In Dretske's terms, the linguistic expression *Cristina writes about cooking* is embedded within a larger expression *Cristina only writes about cooking*. The expression *only* is sensitive to discourse importance (Dretske's contrast), realized by prosodic prominence.

Other phenomena, such as question-answer congruence, are sensitive to discourse importance but their effect is not truth-conditional. In (24-25), only one of the answers is appropriate for the question, even though the two answers are truth-conditionally equivalent. In this sense, discourse importance is used pragmatically, as a means of managing the discourse, rather than semantically to change the literal meaning.

- (24) Who stole the cookie from the cookie jar?
 - a. ADELYN stole the cookie from the cookie jar.
 - b. #Adelyn stole the COOKIE from the cookie jar.
- (25) What did Adelyn steal from the cookie jar?
 - a. #ADELYN stole the cookie from the cookie jar.
 - b. Adelyn stole the COOKIE from the cookie jar.

Embedded in a larger expression (cf. 26-27), the effect of discourse importance survives. Following Dretske, the phenomenon of question-answer congruence is also sensitive to discourse importance (contrast) and requires the grammatical "resources" to distinguish such minimal pairs.

- (26) Who stole the cookie from the cookie jar?
 - a. **I highly doubt that** ADELYN stole the cookie from the cookie jar.
 - b. **#I highly doubt that** Adelyn stole the COOKIE from the cookie jar.

- (27) What did Adelyn steal from the cookie jar?
- a. # **I highly doubt that** ADELYN stole the cookie from the cookie jar.
 - b. **I highly doubt that** Adelyn stole the COOKIE from the cookie jar.

A diverse, apparently heterogenous, set of linguistic expressions exhibit sensitivity, either semantic or pragmatic, to discourse importance, including the list in (28) from Beaver & Clark (2009:4). (Following Beaver *et al.* (2007), I will refer to constructions or configurations that are affected by discourse importance as “focus sensitive” and the phenomenon as “focus sensitivity” or, following Jackendoff (1972), “focus association”.)

- (28) **exclusives:** *only, just, merely, ...*;
non-scalar additives: *too, also ...*;
scalar additives: *even*;
particularizers: *in particular, for example, ...*;
intensives: *really, totally, ...*;
quantificational adverbs: *always, usually, ...*;
determiners: *many, most, ...*;
sentential connectives: *because, since, ...*;
counterfactuals: *if it were ...*;
emotives: *regret, be glad, ...*;
superlatives: *-est*;
negation: *not, no, ...*; and
generics: e.g. *Mice eat CHEESE*.

Nor are these focus sensitive constructions and configurations functionally homogenous. Dik (1980), for example, distinguishes six different usages of focus: completive, parallel, replacing, restricting, expanding and selecting.

Can we and should we analyze focus sensitivity as a single phenomenon? One kind of approach to this question has been to locate discourse prominence entirely in pragmatics or semantics; or entirely within or outside of grammar.

An attempt of the first kind involves the notion of domain restriction (also known as free parameter selection; cf. Rooth 1985). Many expressions of natural language quantify over a

context-sensitive domain. For instance, *every semanticist* may refer to one set of individuals in one context and another set of individuals in another (cf. 29-30).

- (29) Every semanticist attended the reception.
(e.g. every semanticist at a particular conference)
- (30) Every semanticist is an avid swimmers.
(e.g. every semanticist at Cornell)

Rooth (1985,1992) and von Stechow (1994) propose that the role of discourse importance is to further constrain a context-sensitive domain. In (31), for example, let us say that *always* quantifies over propositions. In (31a), *always* quantifies over propositions in which people escorted ballerinas; in (31b) *always* quantifies over propositions in which officers escorted people. Note that the two utterances are truth-conditionally distinct: (31a) is false if there was a bank clerk escorted a ballerina, but (31b) would be true.

- (31) In Saint Petersburg, officers always escorted ballerinas.
 - a. OFFICERS always escorted ballerinas.
(i.e. whenever someone escorted a ballerina, it was an officer)
 - b. Officers always escorted BALLERINAS.
(i.e. whenever an officer escorted someone, it was a ballerina)

In addition to the ordinary semantic value of word, phrase or sentence, Rooth (1985) posits a parallel, but fully compositional level of meaning called the “focus semantic value”.¹⁰ Discourse importance constrains the domain of *always* to a subset of the focus semantic value. Roughly, the focus semantic value is the set of propositions, properties, individuals, etc. obtainable by replacing every discourse important part with an alternative of the same type--accordingly, the theory is known as alternative semantics.

Continuing with the same example, suppose that the focus semantic value of *BALLERINAS* is the set {ballerinas, singers, nurses}; the focus semantic value of *escorted*

¹⁰ In Rooth (1985) it is called the “p-set”.

BALLERINAS would thus be the set of properties {escorted ballerinas, escorted singers, escorted nurses}; and the focus semantic value of *Officers escorted BALLERINAS* would thus be the set of propositions {Officers escorted ballerinas, Officers escorted singers, Officers escorted nurses}. In (29b), the domain of *always* is restricted to a subset of the focus semantic value of *Officers escorted BALLERINAS*: i.e. the domain may consist only of propositions having the form ‘Officers escorted x’.

Pragmatically-oriented approaches such as the domain selection analysis have been contrasted with the approach of the Prague school theorists (e.g. Hajičová 1983; Sgall 1984; Hajičová 1984; Sgall *et al.* 1986; Peregrin and Sgall 1986; Materna, Hajičová 1987) and the structured meaning approach (e.g. von Stechow 1985/1989; Krifka 1992). On the latter approaches, the proposition expressed by a sentence is bi-partitioned into a background and a focus. The representation may be semantic or even syntactic and these approaches are therefore often referred to as “grammaticized” or “semantic”, in contrast to “degrammaticized” or “pragmatic” approach like Rooth (1985,1992) and von Fintel (1994).

I am not convinced that the distinction is useful, however. To the extent that the constraint on domain restriction is uniquely linguistic and part of a rule system, it is part of the grammar. And to the extent that the constraint on domain restriction affects truth-conditional meaning and applies to semantic objects (i.e. focus semantic values), it is (at least partly) semantic.

A larger concern, often characterized as competition between semantic and pragmatic approaches or between grammaticized and de-grammatized approaches, is whether we can understand discourse prominence as a general phenomenon or as an idiosyncratic property of individual semantic constructions and discourse configurations: whether our theory of discourse importance is explanatorily “strong” or “weak” (Rooth 1992). The domain restriction approach is a strong theory of discourse importance in the sense that a single mechanism is responsible for all of the sensitivity effects. A theory which requires construction or configuration specific rules, on the other hand, is weak in the sense that it must be stipulated for each semantic construction

or discourse configuration how it interacts with discourse importance. For example, if we must write the semantics of *always* by referring directly to a representation of focus (e.g. the focus semantic value or the focus partition of structured meaning) and write the semantics of *only* by referring—entirely independently—to this same representation of focus, and so forth for each construction, then our theory is weaker than a theory which stipulates a single principal accounting for focus sensitivity in all constructions.

A weak theory is predicted to be deterministic: marking discourse importance is obligatory because a given construction or configuration is directly dependent on discourse importance for its meaning. A strong theory is predicted to be non-deterministic: marking discourse importance is optional because the meaning of a given construction is affected only indirectly by discourse importance.

Beaver & Clark (2008) elaborate an intermediate theory, suggested in Rooth (1992), according to which only some focus sensitive constructions are deterministically sensitive.

Prosodic prominence as a correlate of discourse importance has played a central role in the debate surrounding the optionality of marking discourse importance. If focus effects are optional, then we should find cases in which focus is not marked.

The phenomenon of second occurrence focus—focus which is repeated or contextually given—was initially argued to provide evidence supporting a strong, non-deterministic approach. Impressionistically, Partee (1991) and others observed that the second occurrence of putatively focus-marked constituent lacked prosodic prominence. For example, *only* associates with *graduate students* in (32) and we intuit that it is realized with prosodic prominence; in (32b), *only* again putatively associates with *graduate students*, yet we intuit that *graduate students* is realized with less prominence than in the first occurrence.

- (32) A: Eva only gave xerox copies to the GRADUATE STUDENTS.
B: No, PETR only gave xerox copies to the graduate students
(based on Partee 1991)

Experimental investigation of the prosody of second occurrence focus has confirmed that the second occurrence of a putative focus lacks intonational prominence (e.g. Rooth 1996a, Bartels 2004, Beaver et al. 2007, Bishop 2008, Howell 2010). Nonetheless, experimental studies have also found that second occurrence foci may be realized with statistically significant acoustic correlates of stress (e.g. increased duration and intensity), although it is unclear how reliably second occurrence focus is realized with stress and whether it is regularly perceived and exploited by listeners (cf. Howell 2010).

The question of treating discourse importance as a unitary phenomenon has also been probed with respect to the dimension of novelty/givenness and contrast. Rooth's (1985, 1992) alternative semantics emphasizes contrast, while the structured meanings approach emphasizes novelty/givenness; yet both approaches assume that discourse importance can ostensibly be characterized by a single notional category, focus. Many other theorists have proposed that novelty/givenness and contrast constitute two distinct notional categories. Particularly in the generative tradition, there has been a fairly even balance between so-called "splitters"¹¹ and "lumpers"¹².

As an instance of novelty/givenness, consider the following examples from Ladd (1980, 1996). In the examples (33-35), there appears to be a negative correlation between prosodic prominence and discourse newness. In other words, *French toast*, *German* and *whisky* have already been introduced in the discourse and we intuit that they are also realized with a lack of prosodic prominence.

¹¹ The terms "splitters" and "lumpers" is due to Beaver & Velleman (2011). Kratzer & Selkirk (2009) cite as splitters Halliday (1967), Chafe (1976), Rochemont (1986), Kiss (1998), Vallduví & Vilkuna (1998), d'Imperio (1997), Pierrehumbert & Beckman (1998), Frota (2000), Face (2002), LeGac (2002), Selkirk (2002, 2007), Neeleman and Szendroi (2003), Féry and Samek-Lodovici (2006), Kratzer (2004), Kratzer and Selkirk (2007).

¹² Kratzer & Selkirk (2009) cite as lumpers Bolinger (1961), Jackendoff (1972), Rooth (1985, 1992), Krifka (1992), Schwarzschild (1999), Ladd (1980, 1996), Gussenhoven (1983, 2004), Selkirk (1984, 1996), Kratzer (1991), Büring (2007, 2008).

- (33) A: Why don't you have some French toast?
 B: I don't remember how to MAKE French toast.
- (34) A: I found an article for you in a German journal.
 B: I don't READ German.
- (35) I brought her a bottle of whisky, but it turns out she doesn't LIKE whisky.

Halliday (1967) characterized givenness as “anaphorically recoverable”, which was formalized by Schwarzschild (1999) as a relation of entailment. A constituent (word, phrase, or sentence) is marked as given if and only if it is entailed, in a formally specified manner, by prior discourse. In the simplest case, a word is marked as given because it has already been introduced.

The phenomenon of second occurrence focus has figured in recent “splitting” accounts of focus because it appears to require the resources of both an alternative semantics based on contrast and an entailment semantics based on givenness or discourse-novelty (e.g. Selkirk 2008). In (32B), *graduate students* is marked with focus in order to generate the appropriate focus semantic value and constrain the quantificational domain of *only*; yet *graduate students* is also marked as given because it is entailed by prior discourse.

“Lumpers” Rooth (2006, 2009) and Büring (2008), by contrast, offer proposals which relate the prosody of second occurrence focus to the relative scope of different focus-sensitive operators. As schematized in (36), a focus-sensitive operator semantically embedded under another focus will have a focus associate that is realized with SOF phonology (e.g. lacking a pitch accent), while the associate of the widest-scope focus operator will have regular focus phonology (e.g. with a pitch accent).

- (36) Configurational SOF (adapted from Rooth 2009)
 [... F ... [[... SOF ...] Op2] ...] Op1

In this dissertation, I adopt a lumping approach from Rooth (2008), which combines the insights of an alternative semantics based on contrasting alternatives with an entailment semantics based on anaphoric relations. As in Rooth (1992), one generates a focus semantic value by substituting alternatives for any focus marked constituents. Unlike the domain restriction account, however, the focus semantic value does not constrain a quantificational domain; rather, in the spirit of Schwarzschild (1999), the focus semantic value constrains what may be a possible discourse antecedent.

For instance, the focus semantic value of (37B) will be a set of propositions such as {Heather canoed the Grand River, Catherine canoed the Grand River, Lesley canoed the Grand River}. By existentially quantifying over the focused marked constituent *Heather*, we get a proposition ‘Someone canoed the Grand River’. This proposition is semantically entailed¹³ by A1, but not A2. This accounts for the intuition of incongruence between A2 and B.

- (37) A1: Catherine canoed the Grand River.
 A2: Heather canoed the Humber River.
 B: No, HEATHER canoed the Grand River.

An interesting case is presented by “all new” sentences (e.g. those sentence which would be appropriate answers to the dinner-table question “What happened today?”). I interpret these as being marked for focus as a single large constituent. By existential quantification of a focused sentence, we get the proposition ‘Something happened’, which is trivially entailed by any discourse antecedent. Broad focus of this kind is the least restricted, which is a desirable result since “all new” sentences may be uttered at the beginning of a discourse. I assume that the prosody of these utterances are determined by phonological and syntactic factors discussed in the literature on focus projection (see above).

¹³ There are no instances in which ‘Catherine canoed the Grand River’ is true and ‘Someone canoed the Grand River’ is false.

- (38) A: What happened today?
B1: Adelyn stole a cookie from the cookie jar.
B2: It rained all morning.
B3: Seventy-six trombones led a big parade through the centre of Chilliwack.

Strict entailment is usually evaluated between overtly expressed propositions, yet many instances of focus are entailed by propositions which have not been overtly expressed in the discourse, but are nonetheless highly salient in the context or may be derived from overtly expressed propositions. In the analysis of (18), the utterance “JOHN traveled to the SOUTH of Italy” requires a discourse antecedent of the form ‘x traveled to the y of Italy’. While the proposition ‘My brother traveled to the north of Italy’ is highly salient because of the explicit utterance “My brother traveled to Milan”, the former proposition is never explicitly expressed.

How to identify in a non-arbitrary way those discourse antecedents which are salient but not overt is a non-trivial, but under-researched question.

2.3 *Syntactic mediation*

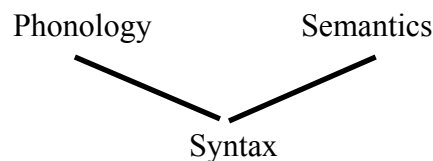
We have considered the phonological, the pragmatic and the semantic mediation of discourse importance and prosodic prominence. There is reason to believe that the correlation is mediated by syntax as well. Empirically, a notion of discourse importance may be signaled by syntactic position or syntactic operation, such as dislocation (cf. 39).

- (39) My father, he's Armenian, and my mother, she's Greek. (Ross 1967)

It is also the case that many languages mark discourse importance morphologically. Whether syntactically or morphologically marked discourse importance belongs in the same category as prosodically-marked discourse importance is worthwhile empirical question and has also influenced many syntactic theories of discourse importance, in particular those positing covert syntactic movement (e.g. Chomsky 1971) and fixed syntactic positions (e.g. Rizzi 1997; Cinque 1999).

Many pragmatically and/or semantically oriented approaches to discourse importance appeal to a syntactic focus feature F, introduced by Jackendoff (1972). The original motivation for this feature was architectural since on the T- or Y-model of generative linguistics the phonological and semantic modules cannot share information and are mediated by syntax. On this view the focus feature F is syntactic, but interpreted phonologically and semantically.

(40) Chomsky's T/Y-model of grammar



2.4 Domains of focus

Many analyses of focus assume, or stipulate explicitly, that the domain or “scope” of semantic focus is restricted to the sentence or semantic proposition. In specifying the relation between semantic focus and discourse prominence, it is further assumed that there is a phonological domain that corresponds to the sentence or proposition. Jackendoff's (1972:237) principle (41) refers explicitly to a focus “of a sentence” and presupposes there is a highest stress in that sentence.

(41) If a phrase P is chosen as the focus of a sentence S, the highest stress in S will be on the syllable of P that is assigned highest stress by the regular stress rules.

While many semantic theories may accommodate a scope of focus smaller than a proposition, Rooth (1992) was perhaps the first to insist on this possibility and to build a notion of scope of focus directly into his semantic framework. Rooth posits a focus interpretation operator \sim^{14} , which fixes the scope of focus and determines a discourse antecedent. (42) is an often cited example, in which the scope of focus is limited to the noun phrase *Canadian farmer*:

¹⁴ Sometimes referred to as “squiggle”.

- (42) a. An American farmer was talking to a CANADIAN farmer.
 b. [An [American farmer]₃ was talking to a [CANADIAN farmer]_F ~ 3]

On an anaphoric theory of focus, *CANADIAN farmer* is licensed by the discourse antecedent *American farmer*; the relation is mediated by the ~ operator and co-indexation.

Truckenbrodt (1995) proposed a correlation between Rooth's scope of focus and the prosodic domains of metrical phonology. We can state this correlation as follows:

- (43) *Stress-F* (based on Rooth 2009)
 Let β be an F-marked phrase with scope ϕ . Then the most prominent syllable in the phonological domain of ϕ falls within the realization of β .

On the autosegmental-metrical theory of prosody, prominence at different levels of prosodic structure have different realizations. Prominence within the phonological phrase is intonational, while prominence at the prosodic word is realized with a particular degree of stress. In other words, phonology appears to provide the resources to signal more than one level or scope of focus.

While this is an attractive picture, the semantic argument for scope of focus below the proposition has been ambivalent. Kadmon (2001) demonstrates that is technically possible to account for most alleged cases without assuming that focus may have scope below the proposition. She concludes that sentence-internal contrasts such as Rooth's "farmer" example are perhaps the only compelling cases.

3 *Methodological pluralism*

The foregoing has outlined some of the challenges in identifying the objects of study for an investigation of the correlation (1). The phonetic resources available to produce prosodic prominence are co-opted for many other purposes and any observation concerning speaker

intension will necessarily be indirect or hypothetical. Moreover, the grammatical mediation of this correlation is evident at multiple levels and obscures any straightforward mapping.

I take for granted, then, that a thorough investigation of this correlation (and indeed any linguistic phenomenon) will require the use of different methodologies, at different times in the course of an investigation, and in iteration. As Borsley (2005) plainly states, “In linguistics as in any other field, what sort of data one requires depends on what questions one is investigating”. This seemingly uncontroversial assumption has not always been shared among scholars of language. Much of modern linguistic theory, for example, has been built on the informal collection of introspective judgments about linguistic examples constructed by the theorist. Many kinds of evidence, in particular “negative evidence” (e.g. intuitions of ungrammaticality or infelicity) about which linguistic constructions cannot exist in a grammar, could arguably not have been obtained at all, or not as cheaply and efficiently, as other methods. Chomsky (1972) famously took the strong position that experimental laboratory methodologies, of the type commonly used in psychology, were simply unnecessary:

The gathering of data is informal; there has been very little use of experimental approaches (outside of phonetics) or of complex techniques of data collection and data analysis of a sort that can easily be devised, and that are widely used in the behavioral sciences. The arguments in favor of this informal procedure seem to me quite compelling; basically, they turn on the realization that for the theoretical problems that seem most critical today, it is not at all difficult to obtain a mass of crucial data without use of such techniques. Consequently, linguistic work, at what I believe to be its best, lacks many of the features of the behavioral sciences. (165)

I have no doubt that it would be possible to devise operational and experimental procedures that could replace the reliance on introspection with little loss, but it seems to me that in the present state of the field, this would simply be a waste of time and energy. (81)

Similar views were espoused against the study of naturally occurring examples in corpora. Robert Lees reportedly quipped to Brown Corpus collaborator W. Nelson Francis in 1962 that his efforts were “a complete waste of time and the government’s money. You are a native speaker of English; in ten minutes you can produce more illustrations of any point in English grammar than you will find in many millions of words of random text” (cf. Biber & Finegan 1991).

Some practitioners of other methodologies approached the “informal” methods of Chomskyan linguistics with equal approbation. Derwing (1979) regards the “blatantly informal” methods of linguists as necessarily “inferior” and leading to “soft data”; Ohala (1975) contrasts the “facile inventions of taxonomic linguists” with “what scientists of language have proven and demonstrated empirically about the behavior of speech sounds”¹⁵; while Becker (1975) rejects all together the “terse, unlikely” constructed examples of linguists:

These example sentences bear no discernible resemblance to the sentences which compose the text that purportedly explains them—yet the linguist’s own sentences are also alleged (implicitly) to be drawn from the same English Language! “

While such attitudes undoubtedly persist in some quarters (e.g. Sampson 2001), the field of linguistics has in general turned towards a more nuanced and pluralistic view of linguistic data and research methods. As Schütze (2005) argues, we must recognize that there is a trade-off in gathering linguistic evidence, between more direct methods like eliciting introspective judgments of grammaticality, similarity or constituency and more indirect methods like naming, phoneme monitoring, click location or truth value judgment.

¹⁵ Quoted from Ladd (1996:13)

On the one hand, we can use tasks for which we have a relatively clear understanding of what our experimental participants [and language consultants, J.H.] are supposed to do, and a relatively justified belief that they share this understanding and are indeed carrying out what is asked of them. [...] These tasks tend to have the property that they are somewhat removed from normal language use; their interpretation is hence rather indirect. [...] On the other hand, we can choose tasks that *prima facie* address our questions of interest much more directly. [...] But then we seem to lose on the other dimension: these tasks generally seem susceptible to the possibility that naïve speakers will not understand what is really being asked of them.

Similarly, Gilquin & Gries (2009) assert the complementarity of corpora and experiments in order to: “(i) solve problems that would be encountered if one employed one type of data only and (ii) approach phenomena from a multiplicity of perspectives”.

The turn towards methodological pluralism in linguistics became increasingly evident in the 1990s. Several research perspectives emerged in which the data, theory and tools from different scientific subcultures were brought to bear on particular areas of linguistic study. Laboratory Phonology (cf. Kingston & Beckman 1990, Cohn *et al.* 2011) brought together researchers in phonology, phonetics, sociolinguistics, language acquisition, speech science and psycholinguistics; Experimental Pragmatics drew together linguists, language philosophers and psychologists (cf. Sperber & Noveck 2004); the Association for Logic, Language and Information (FoLLI) was formed in 1991 to “advance the practising of research and education on the interfaces between Logic, Linguistics, Computer Science and Cognitive Science and related disciplines” (FoLLI website). Syntacticians began to explore the use of experimental techniques from psychology and psycholinguistics (e.g. Schütze 1993, Bard *et al.* 1996, Cowart 1997, Keller 1998) in a perspective that came to be known as Experimental Syntax. The growth of linguistic research using electronic text corpora, born with the work of Francis & Kucera (1964), increased by several orders of magnitude following to the introduction of the scanner

and the increase of typesetting (Kepser & Rice 2005:2, Bonelli & Sinclair 2006:208). And many of the most widely used speech corpora were also published during this decade (examples discussed below; see also Cole & Hasegawa-Johnson 2011).

By about 1990 linguistics had changed from a subject that was constrained by a scarcity of data to one that is confused by more data than the methodologies can cope with. Some may even claim that it has not yet come to terms with this cornucopia (Bonelli & Sinclair 2006:208)

The methodological approach in this dissertation is a product of this general trend. Because it is new, I place particular emphasis on the use web-harvested speech data, yet I am careful not to privilege this kind of data, nor do I believe that there is a single “correct” approach to the use of web-harvested speech. In Chapter 3, I use the web-harvested speech data as many psycholinguists studying syntactic processing or syntactic ambiguity resolution use written corpora: with an explicitly formulated hypothesis and accompanied by cross-validation with artificial but controlled laboratory experiments (Gilquin & Gries 2009). In Chapter 4, I appeal largely to introspective data and theoretical argumentation used in theoretical linguistics. In Chapter 5, I adopt an exploratory and observational approach more commonly found in corpus linguistics (Gilquin & Gries 2009).

4 The value of web-harvested speech

4.1 Limits of creativity: Non-discovery, false discovery

In the remainder of this chapter, I suggest some of the ways in which web-harvested speech make a unique and complementary contribution to research on meaning and prosody generally, and the correlation (1) in particular.

Generative linguists have long been in the business of constructing example sentences, and experimentalists the business of creating sentences as stimuli. In the domain of fiction—e.g.

novels, plays, television and movies—certain writers may be praised for their skill in generating plausible, naturalistic dialogue. Are linguistic scholars similarly skilled, by virtue of their expertise? I suspect that indeed some are, but I will suggest that we be weary of relying too much on data generated in this fashion.

Introspection and laboratory experimentation are at the same time enhanced and limited by the experimenter's creativity. When a researcher intuitively constructs a linguistic construction to have a particular prosodic realization and devises constructed examples or designs an experiment to elicit utterances of the construction, we may reasonably ask to what extent have the data been influenced by her intuitions.

In Chapter 5, I harvest from the web naturally occurring utterances of the expression *he himself*, an adnominal emphatic reflexive (ER; e.g. *himself* in (44)).

(44) Jane met Chomsky himself.

A particular approach to the semantics of the adnominal ER (the FOCUSED ASSERTION OF IDENTITY approach; e.g. Eckhardt 2002, Hole 2002, Gast 2006) is chiefly motivated by the impressionistic observation that the reflexive *himself* is realized invariantly with prosodic prominence. The reflexive is argued, therefore, to be invariantly focused. When focused, the adnominal ER contributes to the meaning of an utterance by evoking alternative relations. In (45), for example, it must be salient in the discourse that Jane met someone standing in a relation to Chomsky, such as his wife or his assistant.¹⁶

(45) [Jane met Chomsky [*himself*]_F] ~

The other leading approach (the FOCUS SENSITIVE OPERATOR approach; e.g. König 1991, Siemund 2000, Bergeton 2004) holds that the reflexive is rather a focus-sensitive operator. In (46), the use of *himself* presupposes that *Chomsky* is ranked highly relative to alternative

¹⁶ I adopt here the convention of syntactic F-marking.

individuals. Since focus-marked *associates* typically receive prosodic prominence rather than focus-sensitive *operators*, one would expect *Chomsky* to have prosodic prominence rather than *himself*. On this approach, the apparent prosodic prominence on the reflexive is a theoretically awkward, idiosyncratic property of certain focus sensitive operators.

(46) Jane met [[Chomsky]_F himself ~]

In the only previous phonological study of the adnominal ER, Ahn (2009) elicits utterances of in the laboratory and finds that the adnominal ER is indeed consistently realized with prosodic prominence (*viz.* pitch accents), which he attributes to its inherent contrastive interpretation.

Careful analysis of the web-harvested data (cf. Chapter 5) reveals that prosodic prominence and semantic focus do occur on *himself* in some instances, but that many other configurations also obtain in which *himself* is not focused or realized with prosodic prominence. For example, many sentences containing the adnominal ER are realized with broad focus on the clause (“all-new” focus) and a prosody of descending *f0* peaks (Figure 3). In this utterance, the target sentence in (47) is presented as new; it need not be salient in the discourse that someone is a really valiant intractable individualist.

(47) He resisted being a member of the party. But he was tucked into the machinery. He was present at Lenin’s death. And he was trying to make it work. He kept believing that it was possible to reinvent society.
[But **he himself** was always a really valiant intractable individualist]
and he began to be criticized for that and vilified for it and increasingly became estranged from his contemporaries.

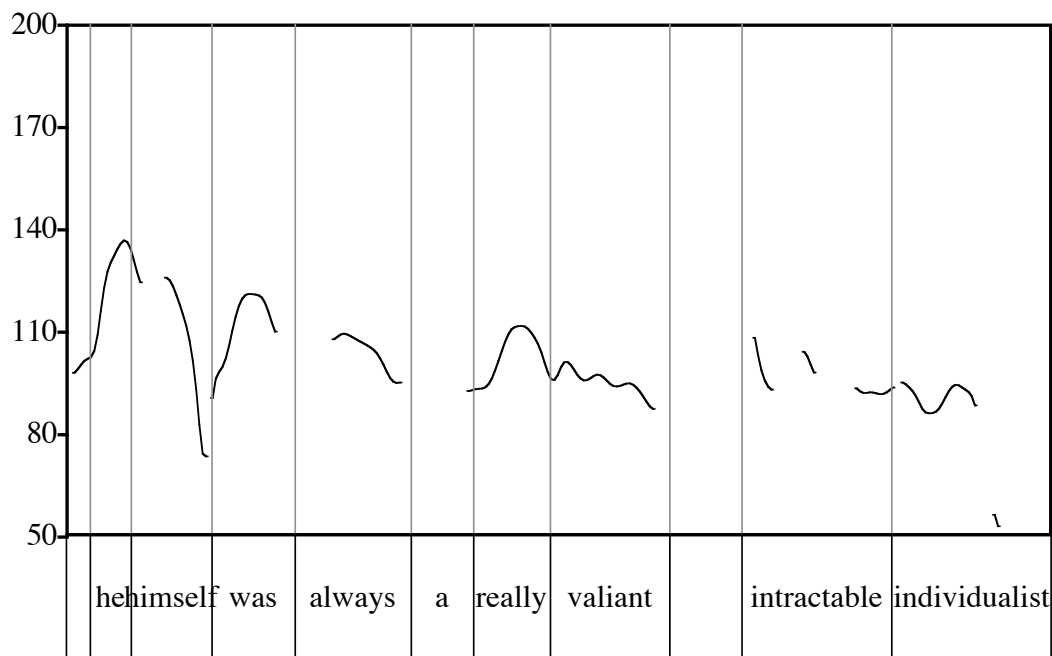


Figure 3. Smoothed *f0* track of (45). Clausal focus configuration.

Constructed examples in the literature on the adnominal ER are typically offered out of context (e.g. 48,49). But as Bresnan (2006) warns, “introspective judgments about decontextualized examples may underestimate the space of grammatical possibility.” In the absence of an explicit context for an utterance, the listener must accommodate one, and there is no guarantee that she will consider all the logical possibilities. I speculate that it simply did not occur to those theorists constructing examples of the adnominal ER that it may be used in a sentence with other focus configurations, such as on the focus on the clause in an “all new” context.

(48) The chancellor himself was surprised at the results. (Siemund 2001)

(49) The president himself gave the order. (König 1991)

And although Ahn takes care to contextualize the target sentences of his experiment in a naturalistic dialogue, his examples are nonetheless constructed by him, and so are influenced

consciously or otherwise by his particular expectations. All of his written scripts for the adnominal ER contain discourse antecedents which license narrow focus on the adnominal ER, precisely the configuration he predicts is obligatory. In his script (50), narrow focus on *herself* in *Jane herself* is licensed by *Jane's mother*. In script (51), narrow focus on *himself* in *he himself* (=‘Perry himself’) is licensed by *Perry's bike*.

- (50) A: Have you seen Jane in the past few years?
 B: No, why?
 A: Well, you know how her mother didn't lose much height in her old age?
 B: Mhm.
 A: **Jane herself has shrunk quite a bit already.**
 B: How much shorter has
- (51) A: Did you hear about Perry?
 B: Yeah – about his bike, right?
 A: Well not only did his bike get hit by a car last week...
 B: Oh no, what happened now?
 A: **He himself was hit just last night.**
 B: Is he okay?
 A: Yeah, the car wasn't going very fast. (Ahn 2009)

This, I believe, is a good opportunity for methodological pluralism. Theory, informed by introspective data predicts a particular correlation between prosody and meaning of a particular construction, and we wish verify the facts experimentally. In setting up the linguistic or non-linguistic context in order to elicit the phenomenon of interest (through scripted dialogue or even unscripted prompts), we are necessarily constrained by our own creativity and influenced to some degree by our expectation as researchers. Corpus-based stimuli, whether presented unaltered or minimally modified for the experimental task can help to mitigate possible biases. In Chapter 3, I follow this tact; I use minimally edited transcriptions of web-harvested data as stimuli for a laboratory experiment of focus in comparative constructions. The adnominal ER merit similar experimental investigation; I leave this for future research.

4.2 *Data quantity, data quality*

Not all corpora contain data of sufficient quality and quantity for studying the correlation between meaning and prosody. Written corpora can reveal effects of prosody through orthographic representations, such as commas or italicization. Beaver & Clark (2008) interpret capitalization as a representation of prosodic prominence in written data from the web. Naturally, this is an imperfect source of data; since marking prominence is not prescriptively required and there are no agreed on conventions, writers do not indicate prosodic prominence consistently. The capitalization in (52-53) may well indicate focus in the scope of the intensifiers *totally* and *fuckin*, but it is also used for the acronyms MTV and OMG. Similarly, the writer of (52) also appears to indicate prosodic prominence by representing the vowel in *way* with an additional character.

(52) MTV like totally gave us TWO episodes back to back! It was like so random. The more the merrier, but it's like waay too much for one recap.

(53) OMG .. this fuckin video is hilarious .. everything about it ,, him being a drunk fuck , and tryn to resist ,, and then how he fuckin falls like a LOG .. omg
(Beaver & Clark 2008)

Material like (52-53) does however serve the purpose of broadening the empirical domain beyond the creativity of the investigator; one can subject written examples to the usual tests involving personal introspection and/or use them as experimental stimuli.

Traditional speech corpora are suitable for many research questions, as well. For investigating the correlation between repeated mention on prosodic prominence, for example, Cole & Hasegawa-Johnson (in press) recommend the Buckeye Corpus (Pitt *et al.* 2007), extended interviews with 40 speakers from Ohio; the Switchboard corpus (Godfrey, Holliman & McDaniel 1992; Godfrey & Holliman 1997), 2430 conversations between speakers of different American dialects on particular topics; the CallHome corpus (Canavan, Graff & Zipperlen 1997),

120 telephone calls between speakers of North American English; or the HCRC Map Task Corpus (HCRC Map Task Corpus 1993), 128 two-person conversations, scripted but conducted under controlled laboratory settings, between students at the University of Glasgow. These corpora are publicly available (although not all are free), and often contain detailed metadata (such as speaker demographics) and rich syntactic or pragmatic annotation (e.g. Ostendorf, Price & Shattuck-Hufnagel 1996, Calhoun *et al.* 2005).

Yet as reported in Chapter 2, these speech corpora often do not contain enough instances of a particular construction of interest, *viz.* a focus-sensitive construction. The Switchboard corpus contains over 240 hours of speech and about 3 million words, yet it contains no occurrences of the expression *he himself*. As for the comparative construction investigated in Chapter 3, a search of Switchboard reveals only 22 instances of the expression *than I did*.

It should also be emphasized that, just as we would not conclude from a search of Switchboard that English speakers never use the expression *he himself*, we should not conclude from the absence of a particular configuration of focus and prosodic prominence that it does not occur. In my web-harvested corpus *he himself*, I identify 7 distinct focus configurations. Two additional configurations are unattested in the corpus, yet I am able to construct plausible examples which illustrate them: I predict that it is possible for *he* to be narrowly focused with scope of focus at the clause level, as illustrated in (54-55). In other words, it is necessary to supplement the naturally occurring corpus data with introspective data. These data alone do not prove conclusively the acceptability of the hypothesized focus configuration; yet they may inform both theory development and experimental design.

- (54) a. The mayor himself will chair the committee.
 b. No, [[he]_F himself will chair the committee] ~
 [uttered while pointing at the Provost]
- (55) a. Who himself will chair the committee?
 b. [[he]_F himself will chair the committee] ~

Provided one has the necessary resources, laboratory experimentation can in principle yield large quantities of controlled data. Yet, as Schütze (2005) notes, we can never be entirely confident that the participants have fully understood the experimental task. This is particularly acute for prosodic phenomena, which Hayes (1995:9) has described as “among the least accessible to consciousness”. We are also generalizing the behavior of a particular demographic (i.e. university undergraduates from a particular geographic location) to a larger speech community.

In Chapter 3, I elicit utterances containing the sequence “than I did” from participants in a laboratory experiment. Results of machine learning classification reveal a robust correlation between certain acoustic measurements and theoretical predictions for location of focus. Although the relationship is statistically significant, it is not exceptionless. In the absence of other kinds of data, we may be left with lingering doubts about whether the results generalize to a larger population of speakers and to natural contexts, and doubts about those instances in which the relationship does not hold. The web-harvest methodology provides complementary data which is largely unscripted, from natural contexts and is drawn from a more diverse population of speakers. The machine learning classification of these data achieves a similar level of accuracy, validating the results from the laboratory data and suggesting that misclassification is not due uniquely to properties of the data source.

CHAPTER 2

WEB HARVEST METHOD

1 *Introduction*

This chapter details the specific methodology¹ I use for harvesting speech from the web. The goal of the methodology is to create databases of multiple repetitions of tokens embedding a fixed word string $w_1 \dots w_n$, within which intonation varies in a way that correlates with syntax, semantics, and/or pragmatics. For instance, in comparative sentences such as (1a,b,c), we have an intuition that contrastive focus in the *than*-clause co-varies with the main clause in a systematic way. A generalization which turns out to be very robust (cf. Chapter 3) is that when reference varies in the subject position between the main and *than*-clauses as in (1a), the subject pronoun *I* in the *than*-clause is prosodically prominent. When reference is constant in the subject position as in (1b) and (1c), the subject in the *than*-clause is not.

- (1) a. She did more than I did.
- b. I wish I had done more than I did.
- c. I did more than I did last time.

The target sequence w_1, w_2, w_3 in this case is “than I did”. In sentences (1a-c), this substring is constant, but intonation varies in a way that correlates with the grammatical context. (1a,b) is a minimal intonational pair, where arguably a single parameter distinguishes the clauses [than I did] in the two utterances. As articulated in formal semantic theories of focus such as Rooth (1992) and Schwarzschild (1999), and accounts of the phonology-phonetics of focus such as Truckenbrodt (1995) and Féry & Samek-Lodovici (2006), this is a parameter which has both a semantic/pragmatic and phonological/ phonetic interpretation.

Constructing indexed web corpora in which such pairs could be retrieved, or collecting large samples of given minimal pairs from web sources, allows the semantic/pragmatic

¹ I developed the web harvest methodology in collaboration with Mats Rooth. A modified version of this chapter has been published as Howell & Rooth (2010).

conditioning of the intonation and its phonetic realization to be studied and modeled on an unprecedented scale. Linguistic theories of intonation ultimately capture correlations between acoustic form and syntax, semantics and pragmatics; they make predictions about what prosodic patterns fit into what grammatical and pragmatic contexts. With this methodology, we can confront deep, logically formalized theories of this correlation with large amounts of data harvested on the web. For some linguistic constructions, it is possible to create datasets of hundreds of tokens; for other constructions, even a hundred tokens is sufficient for some research purposes and several orders of magnitude larger than the data available in traditional, curated corpora.

In Section 2, I detail the particular resources and tools used to harvest the data in this dissertation.² Section 3 evaluates the efficacy of the retrieval using this methodology, discussing sources of error, such as failure to retrieve an audio file over the network, and speech recognition errors. Section 4 offers some conclusions and suggestions about the form of web corpora of spoken language data that would be suitable for research on intonation.

2 *Web harvest method*

I used an external search engine with indexing based on automatic speech recognition to identify the URLs of audio files that contain (or may contain) tokens of the target word sequence $w_1 \dots w_n$. I aimed to use a basic approach of downloading html pages from the search engine, using simple text processing to extract URLs of audio files and other relevant information, retrieving and cutting audio files with software with a command-line interface, and using makefiles and glue languages to control the retrieval and integrate the software components.

² For the most recent developments, the reader is directed to my website (currently <http://conf.ling.cornell.edu/jah238/>) and the website of my collaborator Mats Rooth (<http://conf.ling.cornell.edu/mr249/>).

Kohler *et al.* (2008), which discusses technology and applications for retrieval of spontaneous conversational speech, lists online search engines that index spoken language. My survey indicated that Everyzing (search.everyzing.com) is suitable for the comparatives experiment (Chapter 3) in the following respects:

- (i) Searches for word strings are possible in the query language, including strings involving frequent words (stop words).
- (ii) Initial experimentation indicated that enough data is indexed to retrieve hundreds or thousands of tokens of the strings we are interested in.
- (iii) The indexed material includes a large amount of conversational data, where intonational phenomena of interest are common, and utterances are produced naturalistically.
- (iv) In addition to the URL of an audio file, the search engine returns time offsets for each target word. This makes it possible to automate cutting the audio files.
- (v) Initial experimentation indicated that, for target strings of interest, the accuracy of the engine's speech recognition was good.

Everyzing indexes both pure audio files and files with combined video and audio. Since the size of the files to be retrieved was an issue, I restricted the experiment to audio files to minimize file size. These audio files are always in mp3 format.

I first queried the engine in a browser, in order to determine whether a given string is common enough. After this, the retrieval is performed under program control, in a sequence that mimics what a human would do in interacting with the engine through a web browser.

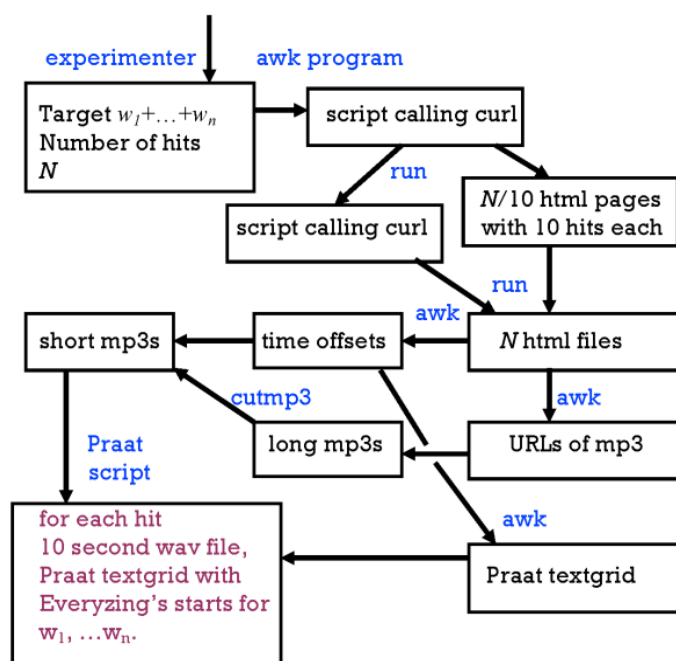


Figure 1. Workflow for mp3 retrieval and editing.

For retrieving material from the search engine, we used curl 7.16.3, which is a command line tool that retrieves data designated in URL syntax (Stenberg 2008). The inputs to the procedure, which is diagrammed in Figure 1, are the target string and the number N of hits to be retrieved.

The first programmatic step constructs a shell program which contains $N/10$ calls to curl. Each involves a URL that embeds the target word string in the format “ $w_1 + \dots + w_n$ ” and an integer which functions as an index into the sequence of hits. Such a string is equivalent to the URL of the page that Everyzing displays when asked in the browser to display a group of 10 hits. Running the shell scripts retrieves $N/10$ html files, each representing 10 hits, and writes another shell script used in the next step. That script calls curl N times, retrieving html files for individual hits. At this point, processing with awk extracts from each file the URL of an mp3, and time offsets for the individual target words in the audio file.

Audio files are retrieved with curl, and subsequent cutmp3, a command line program for cutting mp3 files, is used to cut a 10-second audio file from each long MP3 file, referring to the time offset (Puchalla 2008).

Finally, I prepared data for analysis in the phonetic software package Praat (Boersma 2001). MP3 files were converted to wav format, and using the time offsets of the target words, a Praat TextGrid file was prepared, which aligns the acoustic signal with the target words. Bit rate in the *thanIdid1* dataset varied from 32 to 256 kbits/s and sampling frequency 11025 to 44100 Hz. By comparison, speech files in the often used Switchboard corpus were recorded over the telephone at 8 kbits/s and with a sample rate of 8000 Hz. Note that mp3 is a lossy compression format, which could have an impact on subsequent processing of the audio signal; however these are the available data.

inmyopinion350.hits	html for hits 350-359
inmyopinion360.hits	html for hits 360-369
inmyopinion351.hit	html for hit 351
inmyopinion352.hit	html for hit 352
inmyopinion352.mp3name	URL of audio file
inmyopinion352.cut	time offset for hit 352
inmyopinion352.mp3	long audio file of hit 352
inmyopinion352-b.mp3	10-second audio file of hit 352

Table 1. Files from a retrieval with target “in my opinion”.

In the scripts that issue requests to search.everyzing.com, we used a time delay of 25 seconds between the termination of one curl retrieval and the issuance of the next, to avoid flooding the server. We found that the audio files retrieved from various sources were often very long, and that retrieval of audio files would sometimes hang; therefore we imposed a time limit of 600 seconds for retrieving each audio file.

Files created in a retrieval run for “in my opinion” are exemplified in Table 1. The file inmyopinion352.mp3 is the full audio signal, while in inmyopinion352-b.mp3 signal has been

cut to a 10-second interval flanking a putative occurrence of the target. In the in-my-opinion run the long mp3 files had a median size of 20MB, and a maximal size of 180MB for a two hour and five minute recording of a university forum. The total size of 714 mp3s retrieved in this run is 16.4GB. The run took 24 hours.

Table 2 lists the most common domain names, indicating a predominance of radio content. WEEI, WNYC, KPBS, and WRKO are radio stations; White Rose Society is an archive of progressive radio; the items in the akamai domain comprise three AM radio stations; NPR is National Public Radio. Podtrac is site that matches podcast and advertising content.

116	a1135.g.akamai.net
110	hosted-media.podzinger.com
76	media.weei.podzinger.com
58	feeds.wnyc.org
54	media.libsyn.com
51	podcastdownload.npr.org
50	feeds.feedburner.com
39	library.kraftsportsgroup.com
33	www.whiterosesociety.org
24	www.kpbs.org
21	www.podtrac.com
21	media.wrko.podzinger.com

Table 2. The most frequent domain names in the in-my-opinion run.

3 *Evaluation of retrieval efficacy*

In a pilot experiment conducted prior to full implementation of the procedure described in Section 2, 179 purported tokens of the string “than I did” were downloaded manually by the experimenter via Everyzing and cut manually using Praat. 91 were identified as unique true occurrences of the target.

In one of several subsequent harvests using the procedure described in Section 2, 2,300 tokens of the target string “he himself” (cf. Chapters 5 & 6) were reported by the search engine, and N was set at 300. The shell scripts retrieved 30 html files representing 300 hits, and then retrieved 285 individual hit html files. From these, awk generated 263 files with time-offset information (22 contained no time-offset information). 60 of the 285 mp3 files downloaded were unreadable. Upon further investigation, many of the unreadable files were in fact recoverable by a new search of Everyzing with uniquely identifying text and then manual download. This suggests corruption during the curl retrieval, rather than a corrupt file at the source.

An experimenter listens to all short mp3 files individually and those not containing unique occurrences of the target utterance were rejected. In 16 cases, the cut file contained inaccurate time-offsets, resulting in a short mp3 file that did not contain the purported target. Often this was due to sponsorship information in public radio podcasts which was appended to the mp3 file but did not appear in the Everyzing media player or transcription. In 25 cases, a rejected file contained an incorrectly transcribed token with a near match (e.g. *sees himself*, *um himself*, *eek himself*, *has himself*) or sometimes with nothing resembling the target (e.g. building stuff, purify, independent senator). Four of the short mp3 files were duplicates of previous files. The remaining true, unique tokens of the target numbered 154, roughly one half of the set initially queried. Other retrieval runs yielded comparable, although different results, as summarized in Figure 2.

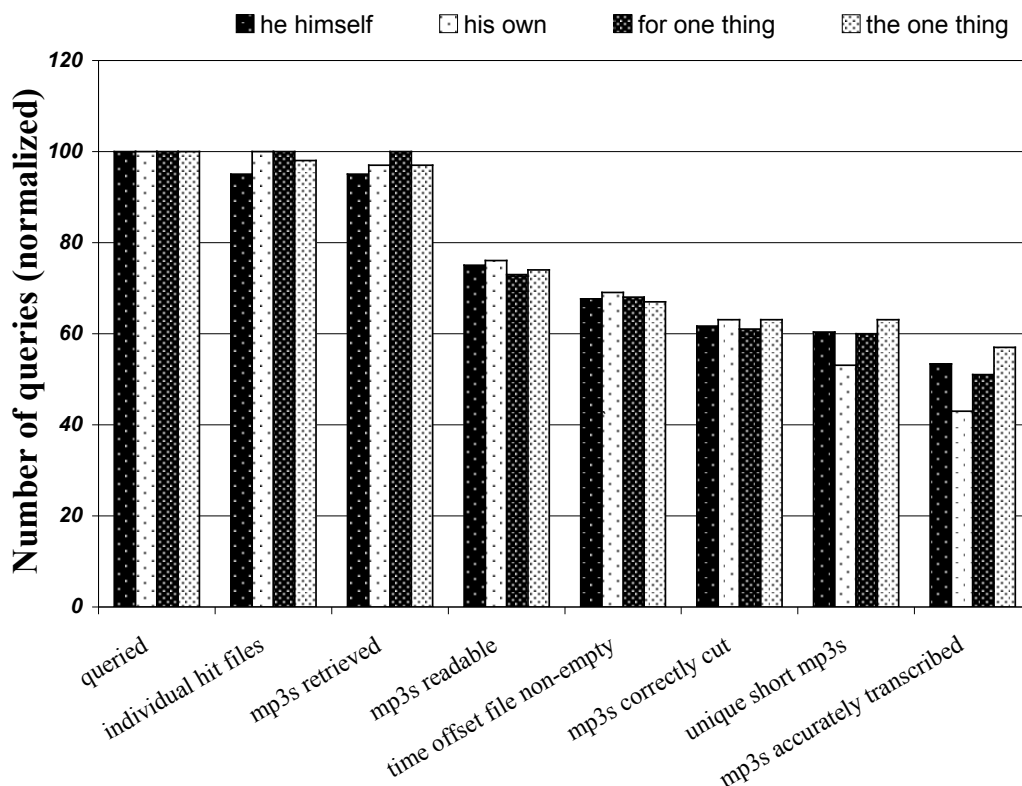


Figure 2. Detailed retrieval efficacy at different processing stages compared for 4 different retrieval runs: (normalized to 100, $n=300, 100, 100, 100$).

I close this section with a comparison of the size of the datasets that can be harvested on the web with a hand-annotated and curated speech corpus. Switchboard (Godfrey & Holliman 1997) contains 240 hours of speech from 2400 telephone conversations, a third of which has been made available by Calhoun et al. (2005) with annotation for syntactic structure as part of the Penn Treebank (Marcus et al., 1993), dialog acts (Shriberg et al. 1998) and information status (Calhoun et al., 2005) and has formed the basis of numerous studies relating prosody, syntax and semantics (cf. Bell et al., 2009; Calhoun, 2006, 2007, 2008; Sridhar et al., 2008, Nenkova & Jurafsky 2007; Jurafsky et al. 1998). Clearly, this type of static, richly annotated corpus offers many virtues, particularly as a standard of comparison.

Unfortunately, the restricted size of such a corpus due to the limitations of human resources means that it is not always large enough to allow statistical analysis of specific linguistic

constructions. The Switchboard-1 corpus available at the Linguistic Data Consortium Online contains 26,151,602 word tokens. Figure 3 compares, for each of five targets, (a) the number of tokens contained in the Switchboard sample (b) the number of true tokens we have already collected and verified from Everyzing, and (c) the projected number of true tokens from Everyzing based on the number of hits returned and assuming a roughly 50% retrieval efficacy. While the Switchboard data may prove a useful baseline for certain target expressions, it is clear that a dynamic web harvested corpus will be not only less costly but much greater in scope. In particular, this allows one to apply machine learning techniques as an alternative to prosodic annotation by human experimenters which necessarily introduces certain theoretical assumptions such as the prosodic ontology of the Tones and Breaks Indices (ToBI) framework (Silverman et al. 1992) for prosodic annotation.

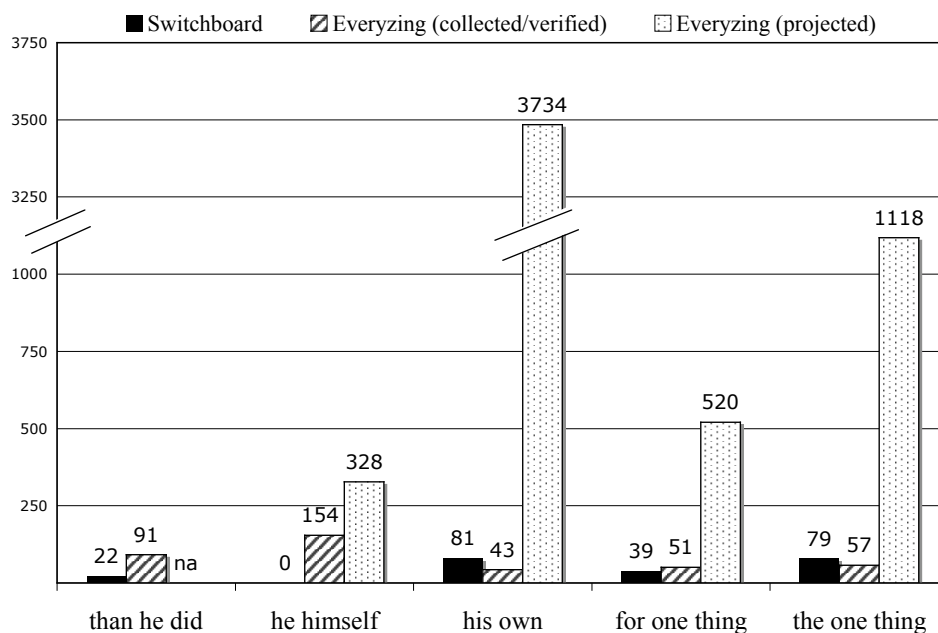


Figure 3. Comparison for each target expression of (a) number of tokens in the Switchboard corpus, (b) number of good tokens already collected and identified in the web-harvested corpus and (c) the number of projected tokens available through Everyzing at the time of harvest, based on total hit count and assuming 50% retrieval efficacy.

4 *Discussion*

In this chapter, I have established by example that large samples of spoken-language phenomena can be gathered on the web using simple web retrieval, text processing, and audio processing methods. The procedure is cheap. Attempted retrieval of 1000 potential tokens results in retrieval of about 750 audio files, containing hundreds of actual tokens of the target. A run of this size requires network transfer and storage of about 20GB of data. Disk capacity for this volume of data costs a few dollars. Network charge environments are readily available where transfer costs for this volume of data is on the same scale. Since the retrieval is done under program control, cost in experimenter time is also small.

The analysis in Section 3 shows that the quality of the retrieved samples varies with the target. Thinking of the system as a prototype concordance interface that presents a list of 10-second audio segments to the linguist for examination, a proportion of 50% of segments that actually contain the target seems acceptable.

Search.everyzing.com went offline in June 2009. Various large sites with indexing bases on speech recognition are online, such as Fox Business News and WNYC. Another interesting angle is provided by individual sites that intend to expose their multimedia material to generic text search by providing transcriptions. For instance, many large broadcast companies, from CBS to PBS, have contracted Ramp to index their audio and video content. Given a list of sites, the tokens can be found with a generic text search engine, or with a textual search engine API.

The current reality is that creating datasets of sufficient size requires interacting with numerous different sites, each with its own HTML representation. Thus the text-processing work that extracts the URL of the mp3 and a time offset would have to be implemented many times, once per site. This could be compensated for by using a more sophisticated scraping technology which works with the Document Object Model representation of the page, rather than simply the string representation like the procedure in Section 3. I hope to look at available systematic solutions to this problem.

A bottleneck in the current procedure is the need for an experimenter to listen to the hits in order to select the actual tokens and create a corrected transcription of the host sentence. This is not really onerous if one is working with a few hundred examples, and at some point we want to evaluate the data as linguists anyway. But suppose 10,000 candidate tokens were available having to listen to about 5000 incorrect tokens just to reject them would be a waste of time. One might build a targeted classifier that, for a single target, attempts to sort out the correct candidates from the incorrect ones. The classifier would be bootstrapped from a manually classified subset. This classification problem is similar to keyword spotting (e.g. Keshet et. al. 2009).

On top of general objections to basing linguistic research on commercial search engines (Kilgariff 2007), in the procedure there are sources of bias in the automatic speech recognition. It seems plausible that a speech recognizer could have substantially different recall rates for two phrase types with the same word string, but different prosodic patterns. If so, the samples collected would be biased in a way that could easily affect the evaluation of linguistic hypotheses.

While it is not possible to avoid the problem within this architecture, one should try to quantify it. This might be done by finding recordings where a correct transcription is independently available. Or if working with a generic search engine, one could put test data onto the web, and measure the recall of the engine for the specific prosodic realizations of the target.

These results and experience are suggestive about suitable forms of indexing for a web corpus of spoken language. As described in Section 3, searches for fixed word strings are useful in finding data bearing on issues on the realization and conditioning of intonation. Such searches appear to compensate for deficiencies in speech-to-text technology, because accuracy at the scale of a short tuple can be good, even if coherent transcriptions are not produced at the sentence scale. Thus it seems attractive to create web corpora of spoken language indexed by

word ngrams, combined with a query system including variables and disjunctions. This would parallel web corpora and concordancing tools for written data (Fletcher 2007).

These results also suggest the feasibility of automatically indexing spoken-language corpora by prosodic features. Assuming that the classification results from Chapter 2 extend to general contexts, an SVM classifier is able to classify tokens of the first person pronoun “I” as focused or not as well as a human, based on local, paradigmatic signal features. This could make it possible to index a corpus automatically with a limited number of prosodic features.

CHAPTER 3

ACOUSTIC CLASSIFICATION OF FOCUS IN A WEB CORPUS OF COMPARATIVES

1 Introduction

In this chapter, the web-harvested data receive a decidedly quantitative, computational treatment. I train machine learning classifiers for the task of focus detection¹, and compare their performance on both web-harvested and laboratory-elicited data. I also compare the performance of the machine learning classifiers to “human classifiers” (i.e. human listeners) in laboratory experiments.

Comparative clauses (viz. *than I did*) were chosen as the targeted phenomenon because they always have an explicit discourse antecedent. The scope of the focus (focus indicated with subscript F) is the *than*-clause, and the antecedent is contained within the main clause.

The location of focus in the comparative clause is thus determined by the matrix clause. On anaphoric theories of focus, we say that the focus “skeleton” (i.e. a proposition with existential closure over the focused constituent) is related anaphorically to the matrix clause. In (1a), ‘he stayed *x* long’ entails ‘there is someone who stayed *x* long’. In (1b), the main clause ‘I liked that song *x* much in the best possible worlds’ entails ‘there are some possible worlds in which I liked that song *x* much’. In (1c), ‘I understand *x* much now’ entails ‘I understand *x* much at some time’.

- (1) a. He stayed longer **than [I]_F did** *He ... I* Class “s”
antecedent: He stayed *x* long
- b. I should have liked that song a lot more **than I [did]_F** *I ... I* Class “ns”
antecedent: I should have liked that song *x* much

¹ There is some variation in terminology between *detector* and *predictor*. I will follow Brenier (2008) in using the term *detector* for a system that uses acoustic features. Brenier also uses the term *detector* for systems using both acoustic and text-based features; the classifiers discussed in this dissertation use acoustic features exclusively. Brenier uses the term *predictor* for systems using only text-based features.

- c. I understand even less **than I did** [yesterday]_F
 antecedent: I understand even *x* little

I ... I Class “ns”

The related² generalization to be tested is that when reference varies in the subject position between the main and than-clauses as in (1a), the subject pronoun *I* in the than-clause is semantically focused and prosodically prominent. When reference is constant in the subject position as in (1b) and (1c), semantic focus and prosodic prominence occur instead on *did* or a following adverbial. We can refer to this generalization as the co-reference criterion (2).

(2) Co-reference criterion for focus in comparative clauses

If the subjects of the main and comparative clauses co-refer, the token belongs to class “s” (subject focus);
 Else, the token belongs to class “ns” (non-subject focus)

With the co-reference criterion, we have an independent way of classifying the comparatives that does not involve prosody. We can train machine learning classifiers to predict the location of focus from acoustic measurements, and we can compare the classifiers’ predictions against the predictions of the co-reference criterion.

This allows us to test the mapping between the speech signal and contrastive focus with naturally-occurring data, while also controlling for semantic/pragmatic context. To a large extent, we also control for syntactic and phonological context, since the string *than I did* is the same in every token.

Variation of other kinds, such as speaker variation, is not controlled for in the web-harvested data, and the extent to which this helps or hinders the automatic classification is an

² Although the co-reference criterion divides instances of the comparative exhaustively, it should be noted that there are certain cases in which it does not correspond exactly with theoretical accounts of focus. In particular, the co-reference criterion does not distinguish cases of double focus, such as (i). The co-reference criterion predicts that (i) belongs in class “s” (subject focus).

- (i) You should have earned less last year **than [I]_F did** [this]_F year
 Antecedent: You should have earned *x* much last year
 ‘You should have earned *x* much last year’ entails ‘someone earned *x* much at some time’.

interesting empirical question. Although some speakers in the corpus data reoccur (e.g. radio and television hosts), most tokens of *than I did* come from unique speakers. In the laboratory experiment, by contrast, each speaker produces several tokens. Individual speakers have different pitch ranges, for example, and may even realize focus with some acoustic cues more than others (cf. Mo 2011). In a traditional laboratory experiment, one models these factors as random effects in a statistical model of the data. It is possible to ask whether some phenomenon occurs only with certain speakers or across all speakers. In the experiments using web-harvested speech, one must necessarily compare across speakers.

Taken together, speech elicited in the laboratory and speech harvested from the web ought to provide the researcher a very close approximation of how speech is naturally produced and perceived under specific conditions. Understanding the limitations of web speech and lab speech are crucial for those instances in which the two provide contradictory results, or in which only one kind of data is available.³

There are a number of parameters by which lab speech or web-harvested speech may potentially be deficient or unrepresentative of natural speech. The web harvest relies on automatic speech recognition which is commercial and proprietary in order to create the indexed transcriptions. We do not know the specifics of the algorithms used in the speech recognition, but it is certain that some tokens will fail to be recognized, and we have no certainty that the missing tokens are randomly or evenly distributed between focus classes (“s” and “ns”), among recording conditions, or among content providers (e.g. different radio shows or networks). As Cole (2011) points out, the speech recognizer may, for instance, perform better on more hyperarticulated speech. And despite the enormous size of the corpus, certain grammatical but low-frequency linguistic conditions are simply unavailable or too difficult to harvest; this was

³ Ideally, one would also include data from traditional, curated corpora; unfortunately, as described in Chapter 2, there are simply not enough tokens of the focus-sensitive constructions we are often interested in.

the case for repetitions of the comparative construction (cf. second occurrence focus) and the comparative construction occurring in questions, to be discussed below.

We also do not know how well the laboratory-elicited data represent speech across speaker demographic, across speech register and across communicative goals. Since the majority of laboratory speech is also “read” speech, it is also possible that they represent different speech genres and that different acoustic features are used in read, laboratory speech compared with less scripted, spontaneous speech.

Given these uncertainties, we want to know to what extent we can use one kind of data to verify the other. In this chapter I ask the follow questions: will a classifier trained on one set of corpus data achieve a similar level of accuracy when applied to new set of corpus data, and when applied to new laboratory-elicited speech? And conversely, will a classifier trained on laboratory-elicited speech achieve a similar level of accuracy when applied to the corpus data?

Another motivation for complementing the web-harvested data with laboratory production data is in order to investigate experimental conditions which do not obtain in the web-harvested data. For this dataset, I am interested in (i) interrogative contexts (viz. questions) and (ii) a phenomenon known as “second occurrence focus”.

Consider first the experimental condition “interrogative”. Typically, a declarative sentence is realized with low f_0 (phonologically a low boundary tone (L%)) at the end of an utterance, and the focus will be realized with a rising-falling f_0 contour (phonologically a high tone pitch accent H*). An interrogative sentence (i.e. a question) is realized with rising f_0 (phonologically, a high boundary tone (H%)) utterance-finally, and the focus will be realized with a falling-rising f_0 contour (phonologically a low tone pitch accent L*).

- (3) H* L%
 Legumes are a good source of vitamins.
 (cf. Legumes are a bad source of vitamins.)

- (4) L^* $H\%$
Legumes are a good source of vitamins?
(i.e. Are you kidding? They're an awful source of vitamins.)

The adjective *good* is contrastively focused in both utterances (3) and (4); however, they are realized with different pitch accents.

A “greatest prominence” theory of focus realization predicts that both intonational and non-intonational acoustic measures may signal the location of focus in these utterances. A “pitch-first” account of focus realization (e.g. Bolinger 1958b, Pierrehumbert & Hirschberg 1990, Steedman 1991, Selkirk 1995a), on the other hand, maintains that the presence of a nuclear pitch accent, whatever its type, is the primary correlate of focus. A nuclear pitch accent effectively entails non-intonational measurements, too, since a nuclear pitch accent must be aligned with the phrasal stress (e.g. Hayes 1995, Ladd 1980, Liberman 1975, Pierrehumbert 1980, Selkirk 1984; in 3' and 4', I represent stress with the metrical grid notation).

- (3')

	H*		L%
	X		
X	X	X	X

Legumes are a good source of vitamins.
(cf. Legumes are a bad source of vitamins.)

- (4') L* H%
- x
- x x x
- Legumes are a good source of vitamins.
- (i.e. Are you kidding? They're an awful source of vitamins.)

Since the type of pitch accent in declarative and interrogative contexts differs, we expect classifiers relying on $f0$ measures specific to one or the other kind of pitch accent to perform poorly across conditions. Since both greatest-prominence and pitch-first theories of focus

realization predict phrasal stress, we expect classifiers using non-intonational measures to perform well across declarative and interrogative conditions.

Second, consider the experimental condition “second occurrence focus” (cf. “secondary accent” Beaver & Velleman 2011), in which focused material is being repeated (cf. 5).

(5) *First occurrence focus (FOF)*

a. At first, you made a very small amount more than [**I**]_{FOF} did.

Second occurrence focus (SOF)

b. Then, after a year or two you made [much]_{FOF} more than [**I**]_{SOF} did.

The phenomenon of second occurrence focus has been central to a debate around the extent to which the focus sensitivity is a property of individual expressions or part of a more general semantic or pragmatic principle. The comparative clause construction is focus-sensitive. If this sensitivity to focus is grammatically encoded, there must be a focused constituent somewhere in its scope in order for the construction to be licit; this predicts that there must be a prosodically most prominent constituent corresponding to the syntactic focus. If, on the other hand, focus in comparative clauses is a purely pragmatic phenomenon, there ought to be instances in which no syntactic focus-marking occurs and hence there is no corresponding prosodic prominence. The repetition of focus-sensitive operators is a testing ground for this production since greatest intonational prominence does not occur on the repeated focus.

A classic example of second occurrence focus comes from Partee (1999). The focus sensitive construction is the exclusive adverb *only*.

(6) A: Eva only gave xerox copies to the [**graduate students**]_{FOF}

B: No, [Petr]_{FOF} only gave xerox copies to the [**graduate students**]_{SOF}

Partee’s impressionistic observation was that in examples like (6) the second occurrence of the putative focused constituent (i.e. *graduate students* in 6b) lacks prominence, and therefore

constitutes a counterexample to a strong, grammatical account which predicts prominence without exception.

Further investigation of the acoustic signal by Rooth (1996a) and Beaver *et al.* (2007)⁴ showed that while such examples indeed lack intonational prominence (*f₀* movement is greatly reduced), the putative focus associate was nonetheless realized with a small but statistically significant amount of increased word duration, intensity and energy, which are phonetic cues of phonological stress. The authors argue that English speakers always realize second occurrence foci with some combination of these acoustic features.

The results of Howell (2010) suggest, however, that speakers do not produce these acoustic measures as consistently as a grammatical account would predict, nor do listeners use the acoustic measures reliably to identify prominence. The stimuli in that study were constructed to be rhythmically even; for many tokens, rhythm had an effect equal in magnitude to focus.

Returning to the present study, grammatically-mediated accounts of focus predict that a robust classifier of FOF ought to perform similarly well on SOF data, especially so if the classifier is trained using acoustic features other than *f₀*, such as duration, intensity and energy.

Sections 2 and 3 discuss the methodologies used for web harvesting and laboratory data-elicitation, respectively. Section 4 introduces the machine learning algorithms—Support Vector Machines (SVMs) and Linear Discriminant Analysis (LDA)—as well as the methods of computer- and experimenter-selected feature selection. Section 5 reports the results of the various machine classifiers and Section 6 reports the results of the human classifiers. The chapter concludes with final discussion and future directions in Section 7.

⁴ See also Bartels (2004), Féry & Ishihara (2009) and Bishop (2008).

2. *Web-harvested data*

2.1 *Web-harvesting*

Two different web-harvested corpora of *than I did* are used in this chapter. The first corpus (*web1*) was collected using the methodology described in the Chapter 2. The tokens were collected using the *Everyzing* search interface from several different content providers. The second corpus (*web2*) was collected with a similar methodology, modified for the (now defunct) search interface *multimedia.play.it* with content from CBS Radio.

Dataset *web1* contained 91 true tokens of the target: 46 tokens with subject focus (s) and 45 tokens with non-subject focus (ns). Dataset *web2* contained 127 true tokens: 62 tokens with subject focus and 65 tokens with non-subject focus.

The antecedent and comparative clause in each token was manually transcribed into English prose. From this transcription, the tokens were manually categorized into one of the two focus categories, according to the co-reference criterion (cf. 2). Although the semantic classification was performed by humans, the task did not require special expertise or training beyond identifying and comparing grammatical subjects of the two clause.

2.2 *Acoustic analysis*

The extraction of acoustic information required annotation at the phonetic level. For each utterance of “*than I did*”, the following phonetic segments were annotated (cf. Figure 1): V1, the vowel [æ] of *than*; N1, the nasal [n] of *than*; V2, the diphthong [aɪ] of *I*; C3, the stop closure and burst of the initial [d] in *did*; and V3, the vowel [ɪ] of *did*.⁵

⁵ Although all material in the string *than I did* may be subject to some degree of variation from the surrounding phonetic/phonological environment, the initial [ð] of *the* and final [d] of *did* were too varied to allow consistent annotation.

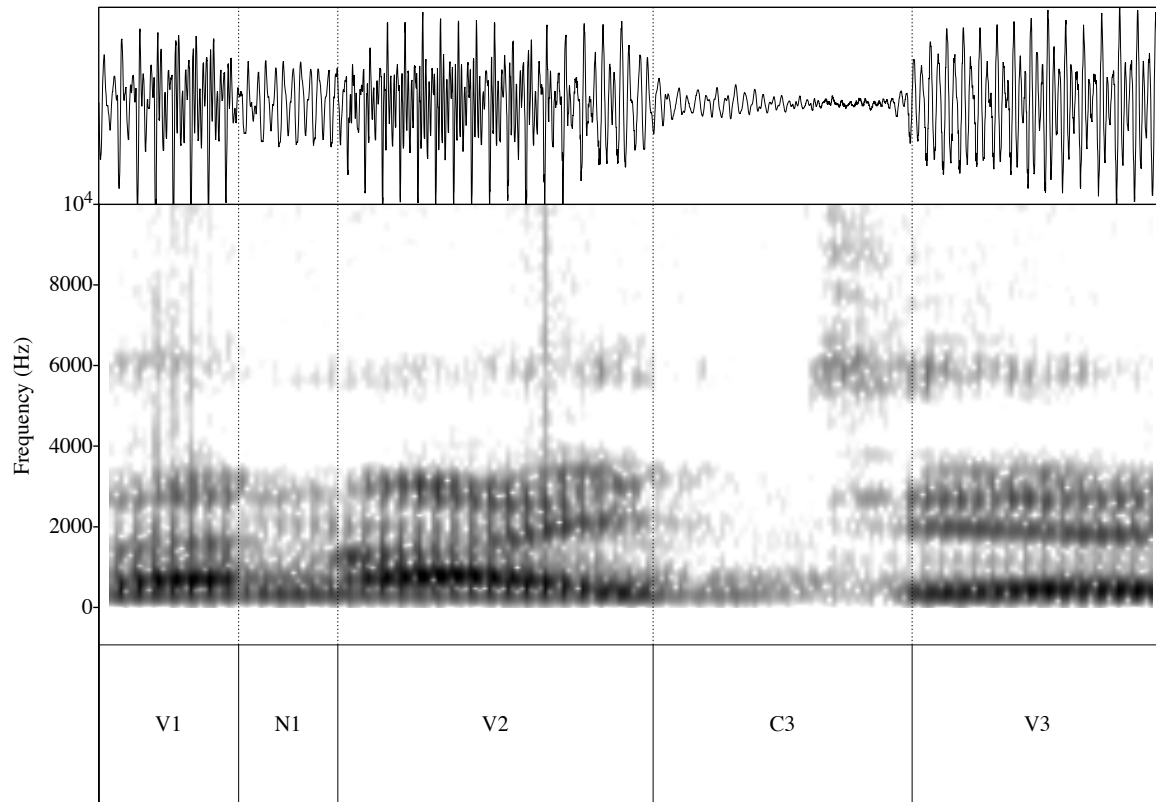


Figure 1. Manual phonetic annotation of a token of “than I did”.

Tokens were labeled manually by the experimenter or by research assistants trained for the task. For segmentation criteria, we used oral and nasal constriction landmarks in the spectrogram and waveform: change in amplitude between vowels and the nasal and oral stops, and the high frequency burst of oral stop releases (cf. Turk et al 2006).

A total of 308 acoustic measures were extracted using the scripting function of Praat (Boersma 2001). Phenomena of interest included duration, fundamental frequency (f_0), first and second formants (f_1 & f_2), intensity, amplitude, voice quality and spectral tilt. Means or extrema were taken for these phenomena, often at various loci such as regular intervals within a vowel or at the time of other extrema. Inter-syllable ratios were also calculated for many measurements,

such as the ratio of vowel length in V2 to vowel length in V3. The full list of measurements is provided with descriptions in Appendix A.

2.3 *Feature grouping: duration, quality, quantity*

In an attempt to corral the 308 features, I have grouped them into related categories. The first group, “duration”, contains all measures encoding only information about timing of the phonetic segments. The second group, “quality”, contains measures related to vowel quality, in particular vocal tract resonances (i.e. formants), and those related to voice quality, including relative amplitude at different harmonics (i.e. spectral tilt or spectral balance) and measures of cycle-to-cycle spectral variation (e.g. jitter and shimmer). The second group, “quantity”, contains measures related to pitch and perceived loudness, including f_0 , amplitude, intensity and power. The groupings carry some degree of arbitrariness; however, no theoretical point hangs on their constituency. They are used simply as convenient starting points for the automatic and manual feature selection discussed in section 4.

3 *Laboratory data*

3.1 *Stimuli*

A total of 16 stimuli were constructed, based on actual tokens from the corpora. Eight of the stimuli contained an ordinary, first occurrence focus and 8 contained both a first occurrence focus and a second occurrence focus. For some SOF sentences, an adjective or verb (e.g. 7) was the FOF and for others a contrasting degree modifier was the FOF (e.g. 8).⁶

- | | | |
|-----|--|--|
| (7) | You worked harder than [I] _{FOF} did,
and I you worked longer than [I] _{SOF} did. | <i>FOF stimulus</i>
<i>SOF stimulus</i> |
| (8) | I think Tom said it a little better than [I] _{FOF} did.
In fact, he said it a lot better than [I] _{SOF} did. | <i>FOF stimulus</i>
<i>SOF stimulus</i> |

⁶ The written stimuli which was presented to participants lacked any annotation for focus (i.e. no capitalization or bracketing).

Among the FOF-only stimuli, half were statements (e.g. 9) and half were questions (e.g.10).

(9) *Declarative context*

There were a lot of photographers who would shoot more than [I]_{FOF} did.

(10) *Interrogative context*

Why do I have more energy today than I did [the day before]_{FOF}?

Each experimental condition was balanced for semantic focus: half “s” and half “ns”. The full set of stimuli are provided in Appendix B.

3.2 *Recording*

Participants were recorded in a sound-attenuated room at the prosody lab at McGill University. Twenty-seven individuals participated, although one participant’s speech failed to be recorded, leaving a total of 26 participants. Each participant was paid.

The stimuli were presented on a computer screen using a set of matlab scripts written by Michael Wagner for conducting prosody experiments, and after reading the text aloud, participants were asked to rate the naturalness of the written stimuli on a scale from 1 (very natural) to 5 (very awkward). The mean rating for the individual stimuli ranged from 1.72 to 3.08; the overall mean was 2.35.

Nineteen tokens were discarded due to disfluencies, such as false starts, hesitations or utterances that did not match the written stimuli, leaving 397 FOF tokens and 195 SOF tokens.

3.3 *Acoustic analysis*

For consistency, the speech from the production experiment was annotated manually into phonetic segments in the same procedure followed for the web corpus data.

The experimental data was also automatically force-aligned, as above. Alignment failed on 3 files for a total of 394 FOF tokens and 194 SOF tokens. Acoustic measures were extracted using Praat.

4 *Statistical classifiers and feature selection*

4.1 *Support vector machines and linear discriminant analysis*

Two machine learning techniques were used to create predictive models of the data. Support vector machines (SVMs) (Boser, Guyon & Vapnik 1992; Cortes & Vapnik 1995) are a relatively recent method of supervised classification that have achieved excellent accuracy in tasks such as object recognition (Evgeniou et al. 2000), cancer morphology identification (Mukherjee et al. 1999) and text categorization (Joachims 1997). Linear discriminant analysis (LDA) (a variant of Fisher’s 1936 linear discriminant analysis) has been used widely for several years in pattern recognition tasks.

For both classifiers, a decision function h of the general form (11) divides the space of attributes into two half spaces according to their labels, in our case “subject focus (s)” or “non-subject focus (ns)” which are mapped to $\{1, -1\}$. In a dataset with only two sets of attributes (e.g. first vowel duration and second vowel duration), the decision function may be represented geometrically as a line dividing a 2-dimensional space (Figure 2), or in a dataset with three sets of attributes, a plane dividing a 3-dimensional space. Generalizing to n -dimensions, we say that the decision function is a hyperplane, an affine subspace of dimension $n - 1$.

(11) $h(x) = \text{sgn}(\langle w \cdot x \rangle + b)$, where

w is a weight vector with direction perpendicular (orthogonal) to the hyperplane
 b is the “bias” which moves the hyperplane parallel to itself
and the decision rule

$$\text{sgn}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$

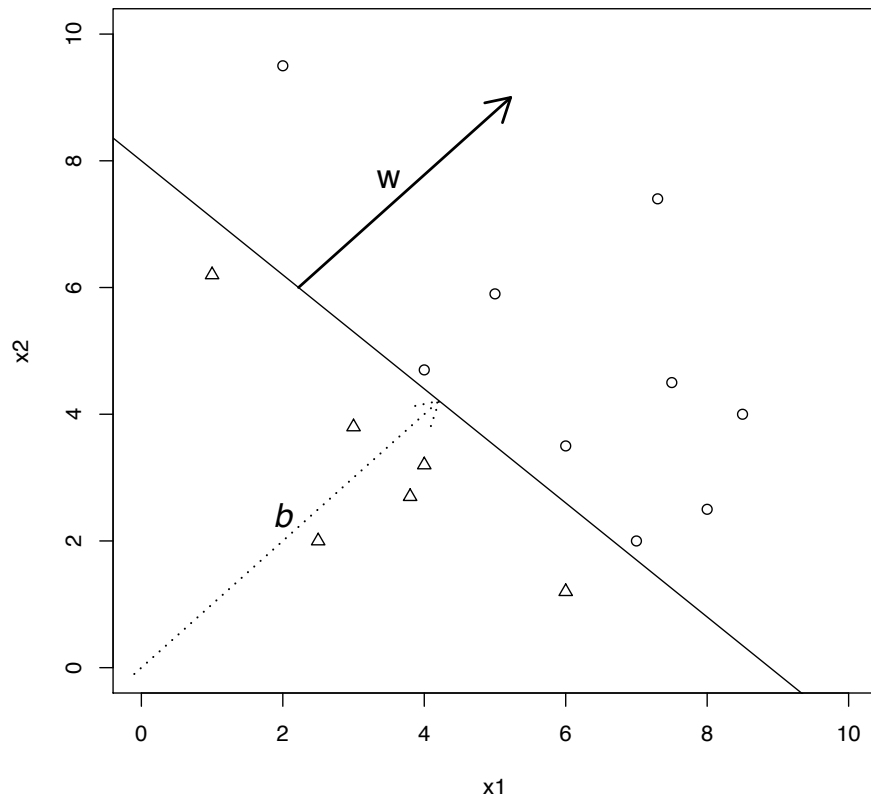


Figure 2. A separating hyperplane in two-dimensional space, with bias b and weight vector w (based on figure in Cristianini & Shawe-Taylor 2000:Section 2).

An LDA classifier looks for the optimal model which minimizes within-class distance and maximizes between-class distance (cf. Figure 3). This approach is considered global, since the optimization is based on the mean and covariance of the classes, which are usually obtained via a discriminant function of ordinary least-squares or maximum likelihood estimation⁷. LDA makes many assumptions, including normal distribution of classes and homogeneity of covariances. Classes in the corpora of this chapter are well balanced, although it is unlikely that the variances

⁷ The R function `lda` in R package MASS is described by its authors in Venables & Ripley 2002 *Modern Applied Statistics with S*, 4th ed., pp. 331-334. To the best of my knowledge, they do not describe the method of optimization.

of all 308 attributes are normally distributed. Poor results may also obtain if the training set is small⁸. Furthermore, the LDA classifier has been shown to perform best when the number of attributes is minimized (ideally no greater than 2 attributes for a binary classifier) and the attributes are not intercorrelated (cf. Brown & Wicker 2000).⁹ In the next section I will discuss a methods of attribute selection, since 308 attributes is a considerable size for the number of examples in the corpora—prescriptively too many—and it is certainly the case that many of the attributes will be correlated. In practice, however, it is often possible to obtain good results even with small datasets and with data which violate the assumptions of normal distribution and homogeneity of covariances (e.g. Lachenbruch 1975; Klecka 1980; Stevens 2002). The implementation of linear discriminant function analysis I used is available in the MASS package (Venables & Ripley 2002) for the statistical computing environment R (R Development Core Team 2008).

⁸ The sample size should be 10 times the number of attributes according to Brown & Tinsley (1987), 20 times the number of attributes according to Stevens (2002).

⁹ The methods of regularized discriminant analysis (Friedman 1989) or shrinkage discriminant analysis (Ahdesmäki & Strimmer 2010) have been proposed to improve performance of simple discriminant analysis when the number of attributes exceeds the size of the dataset. I do not pursue these methods here.

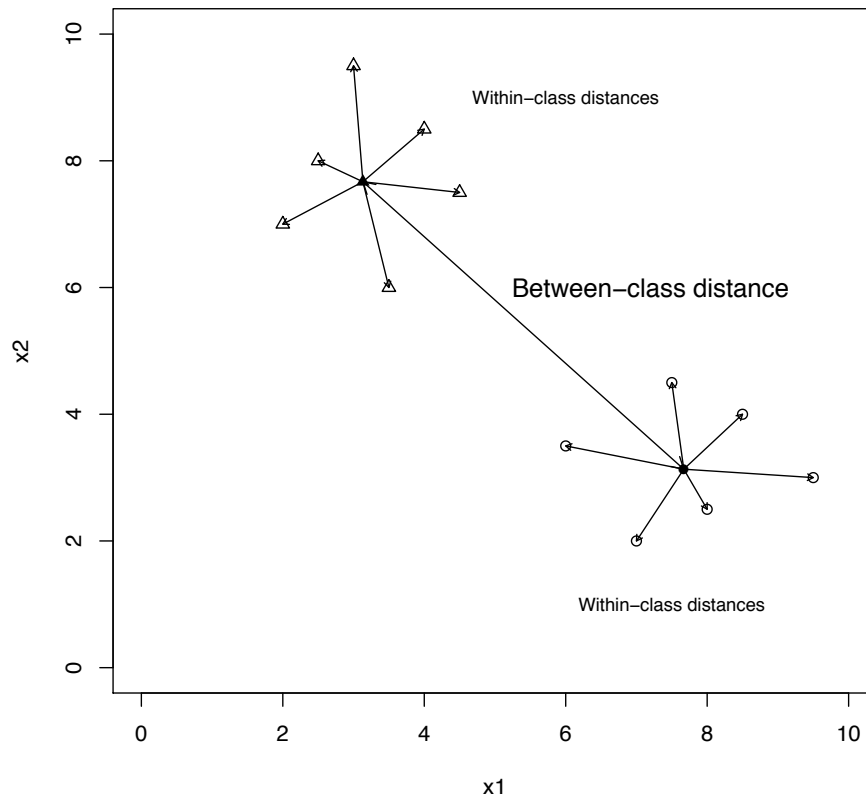


Figure 3. An linear discriminant analysis (LDA) classifier looks for the optimal model which minimizes within-class distance and maximizes between-class distance.

An SVM classifier looks for the optimal model which maximizes the margin between classes. This approach is considered local, since the optimization is based on data at the boundaries between classes (i.e. the “support vectors”). This is illustrated geometrically in Figure 4 for a two-dimensional space. SVM can outperform many conventional classifiers when the number of training data is low and the number of attributes is high. As a maximum margin classifier, SVM also does not assume that the classes are normally distributed or that the classes have equal covariances, although it shares with most classifiers the assumption that the training and test data are independent and produced in the same way.

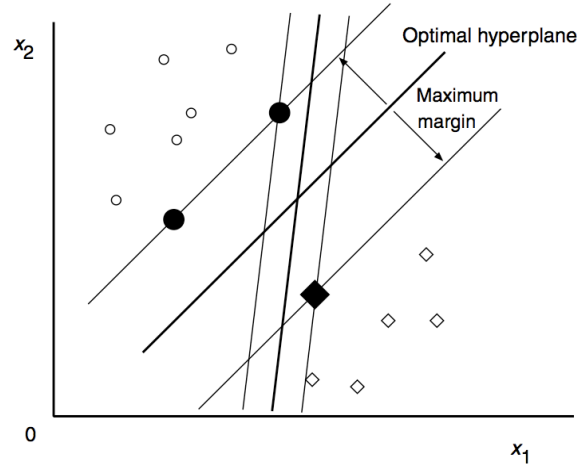


Figure 4. Optimal separating hyperplane in a two-dimensional space (reproduced from Abe 2005:17)

Another feature of SVMs is the mapping of linear attributes into a multi-dimensional feature space¹⁰, the so-called “kernel trick”. By expressing the decision function in dual coordinates, as in (12), it is possible to introduce a kernel function. This greatly reduces the complexity of the algorithm and allows it to scale well with a large number of examples. Although the data should be internally scaled for best results, use of a non-linear kernel also avoids the need to transform attributes which may be non-linear (e.g. duration, energy).

$$(12) \quad h(x) = \text{sgn} \left(\left\langle \sum_{j=1}^n \alpha_j y_j \cdot x \right\rangle + b \right)$$

Many kernel functions have been used successfully in different classification tasks. Hsu *et al.* (2003) recommend a radial basis function (RBF)¹¹, a non-linear mapping which has been shown to also encompass a linear kernel (Keerthi 2003) and behave similarly to a sigmoid kernel (Lin &

¹⁰ The terms ‘feature’ and ‘attribute’ are used here in their statistical or computational sense, referring to a particular vector of data (e.g. the vector of data corresponding to 2nd vowel duration). Note also that the terms ‘feature’ and ‘attribute’ are often used to distinguish predictors before and after kernel mapping, respectively. Since nothing in the study hangs on this distinction, I will use the terms interchangeably.

¹¹ Equation for RBF kernel: $K(x, x^0) = \exp(-\gamma \|x - x^0\|^2)$

Lin 2003). Hsu *et al.* note that the RBF kernel requires only two hyperparameters, while a polynomial kernel, for example, will contain two or more, contributing to model complexity. (All kernels contain at least one hyperparameter C , cost or constant.) At the same time, Hsu *et al.* also suggest that the results of a linear kernel may be comparable with those of an RBF kernel in situations where the number of attributes to be mapped is greater than the number of data instances, a situation which obtains with a full model of the web-harvested dataset. I will therefore consider both RBF and linear kernels. The implementation of SVM I use is available in the *libsvm* package (Chang & Lin 2001) for R.

In a benchmark study, Meyer *et al.* (2003) compared the performance of the SVM (RBF kernel¹²) and LDA classifiers in R on 21 real-world and artificial datasets, and found them to have comparable performance, with SVM outperforming LDA on 19 of the 21 datasets on measures of error rate or hit rate. The authors did not offer an explanation for the better performance of LDA on two of the datasets.

The LDA and SVM classifiers also require datasets without missing values. Algorithms in Praat and other acoustic analysis software have a notoriously difficult time extracting values such as f_0 in the absence of regular, periodic voicing (i.e. normally produced vowels). A dataset with missing values was therefore unavoidable, and many values were undefined. No prescribed method for handling data with missing values exists. I followed the procedure of replacing all undefined values with zero.

4.2 Redundant features and feature selection

In finding a classifier for a given linguistic dataset, we may be concerned simply with the classification task itself: developing and improving the ability of a particular decision function to generalize from one set of data (the training set) to another (the test set). I will call this the “functional measure” (Cristianinni & Shawe-Taylor 2000). We may also be concerned with how

¹² The authors observe in a footnote that “additional experiments showed that by the use of a linear kernel, similar results as with LDA could be achieved” (p.176).

the classification task is achieved and how closely it models real human cognitive ability, the “descriptive measure”. The relative importance of the two measures typically varies according to the goals of the researcher. Consider the following functionally-oriented view from Cristianinni & Shawe-Taylor.

Shifting our goal to generalisation removes the need to view our hypothesis as a correct representation of the true function. [...] In this sense the criterion places no constraints on the size or on the ‘meaning’ of the hypothesis – for the time being these can be considered to be arbitrary. (Cristianinni & Shawe-Taylor 2000: Section 1.2)

Another more descriptively-oriented researcher concerned primarily with the underlying or “true” function may be wary of even a high-accuracy decision function which incorporates what may seem to be linguistically irrelevant or orthogonal noise in the data. In practice, however, the functional and descriptive are not mutually exclusive and are, hopefully, mutually informative.

From a functional standpoint, feature selection (elimination, extraction, reduction, filtering, screening, etc.) for LDA is a necessity where the possibility of collinear features exists. Indeed, the R implementation of LDA is halted and cannot proceed in case of high collinearity. As for SVM, one reason to use this classifier is precisely to avoid costly feature selection; nonetheless, feature selection prior to or in the process of building an SVM classifier has been shown to improve the generalization accuracy and/or model complexity (and thus computation) for those datasets with redundant or irrelevant features. For example, Sarojini (2009) demonstrates improved accuracy for a clinical dataset with a large number of instances (768) and a small number of features (8 prior to feature elimination) while conversely Weston et al. (2001) demonstrate this effect for cancer discrimination in a dataset with a small number of instances (72) and a large number of features (7129 genes).

Descriptively, feature selection is a means of peering into the “black box”, and understanding which features are contributing to a model’s generalization accuracy. For example, a classifier which accurately predicts a focus category may be the goal, but we also wish to know which acoustic measures are important for this task. The set of acoustic measures used by a classifier to predict focus are not necessarily equivalent to the set of acoustic measures that a human speaker uses to convey focus or to the set of features an individual human listener uses to interpret focus; however, the question of whether and why the machine-learning and human sets of attributes are not equivalent is in fact a useful research question provided by the classifier.

Most authors agree that some combination of manual and statistical feature selection techniques may be used, although there is no consensus on the ordering or relative importance of manual or statistical feature selection:

Most authors agree that some combination of manual and statistical feature selection techniques may be used, although there is no consensus on the ordering or relative importance of manual or statistical feature selection:

Feature selection should be viewed as a part of the learning process itself, and should be automated as much as possible. On the other hand, it is a somewhat arbitrary step, which reflects our prior expectations on the underlying target function. (Cristianini & Shawe-Taylor 2000: Chapter 3)

To start, the initial variable list should be logically screened, based on substantive theory, prior research, and reliability of measures, as well as on practical grounds. Next, the list can be statistically screened. (Huberty 2006:11)

Of course, an investigator's professional opinion also can be relied upon when selecting potential discriminator variables. (Brown & Wicker 2000:212)

VarSelRF (Diaz-Uriarte 2009) is a feature selection algorithm designed for genetic research, in which datasets typically contain large sets of features for relatively few data instances. This algorithm is based on a random forests method of classification and uses “backwards variable elimination (for the selection of small sets of non-redundant variables) and selection based on the importance spectrum ([...] for the selection of large, potentially highly-correlated variables)” (from package documentation, Diaz-Uriarte 2009). This is a “filter” method of feature selection, since it occurs as a kind of preprocessing before a model is trained.

As a means of investigator-guided statistical feature selection, I use varSelRF to determine: the “best” set of attributes among all 308 attributes; the best set of measures among “duration” attributes; the best set of measures among “quality” attributes (e.g. formants, spectral tilt, jitter; cf. Section 2.3); and the best set of measures among “quantity” attributes (e.g. f0, amplitude, intensity, power; cf. Section 2.3).

In the case of SVM models, note that the hyperparameters and feature set will influence each other, so each model will in theory require different hyperparameters. Hyperparameters are tuned for a given model in the R implementation of SVM, but pure “filter” feature elimination methods that occur prior to training a model do not take this effect into account. Therefore, I also consider recursive feature elimination (RFE) according to F-score or harmonic mean—the ratio of the between- and within-class variances—with an SVM model tuned to each feature set under consideration. A feature selection tool, written in the programming language Python,¹³ was developed by Chen & Lin (2006). Sarojini (2009) applied this F-score elimination plus SVM to a Pima Indian Diabetes dataset and found that an SVM with a feature set reduced in this way outperformed an SVM model with all attributes. Huang et al. (2008) used F-score elimination for an SVM classifier of breast cancer identification from genetic data and found that it performed slightly better than LDA.

¹³ The algorithm is the same (libSVM) as that used in the R environment.

Finally, I also consider models with experimenter-selected (i.e. hand-picked) feature sets, informed by both the statistical feature selection results above and my theory-informed intuitions.

The methods of feature selection are summarized in (13). The individual features selected (e.g. duration of second vowel, ratio of f0 maxima, etc.) depend on the training data, and are discussed for web-harvested training data in Section 5.2.1 and laboratory training data in Section 5.3.1.

- (13) Methods of feature selection
 - (i) no feature selection*
 - (ii) VarSelRF
 - (a) VarSelRF all features (*)
 - (b) VarSelRF duration features (*)
 - (c) VarSelRF quality features (*)
 - (d) VarSelRF quantity features (*)
 - (e) (c-e) together (*)
 - (iii) hand-picked
 - (iv) F-Score + SVM

*for SVM only; feature set with collinear features incompatible with LDA

5 Machine learning experiments

5.1 Datasets and evaluation of classifier performance

In the previous section, we reviewed several different kinds of classifiers, differing by the particular algorithm used for classification (linear SVM, radial SVM and LDA) and differing in the methods (automatic and manual) of variable selection.

In this section, we begin by reviewing the different datasets used for training and testing of the machine learning classifiers, and the measures I'll use to evaluate them.

The datasets come from the web (Section 2) and from laboratory recordings (Section 3). We also consider data automatically forced-aligned and data hand-annotated data separately. The

forced alignment was performed using scripts written by Kyle Gorman¹⁴ and modified by Michael Wagner for the Hidden Markov Model Toolkit (HTK)¹⁵. The lab data was aligned with both methods: using automatic forced-alignment (*lab_htk*) and hand-annotated (*lab_hand*); at the time of writing, the web data were annotated with only the manual method.

The web data is further divided into *web1* and *web2*, which correspond to the first two iterations of harvesting: *web1* was harvested from *Everyzing* and *web2* was harvested from *play.it*. The combined sets of web data are labeled *web_1+2*.

The lab data is further divided according to experimental condition. The first division is between first occurrence focus (FOF) and second occurrence focus (SOF). The second division is between declarative (declarative) and interrogative (interrogative) contexts.

Comparing every possible data source, experimental condition, classification algorithm, etc. in training and testing would quickly become unmanageable. Fortunately, we can safely ignore several of these combinations. The results for the hand annotated lab data (*lab_htk_FOF*) and the automatically annotated lab data (*lab_hand_FOF*) in the FOF condition did not reveal any large differences, so I considered other experimental conditions only with respect to the automatically annotated data. In addition, the SOF data had only a single experimental condition, declarative.

We arrive, then, at list of 8 different datasets to be considered. Table 1 summarizes each dataset, including the abbreviation used, the source of the dataset (web-harvested or laboratory-elicited), its size (for both focus classes “s” and “ns”) and the method used to annotate the data.

¹⁴ See <http://www.ling.upenn.edu/~kgorman/papers/segmentation/.speechseg.html>. For more recent forced aligners, see the ProsodyLab aligner (Gorman, Howell & Wagner 2011) and the Penn Phonetics Laboratory Forced Aligner (Yuan & Liberman 2008a)

¹⁵ see <http://htk.eng.cam.ac.uk/>

Name	source	details	annotation method	size (ns/s)	baseline accuracy
<i>web1</i>	web-harvested	Source: <i>Everyzing</i>	manual	91 (45/46)	50.5
<i>web2</i>	web-harvested	Source: <i>play.it</i>	manual	127 (65/62)	51.2
<i>web_1+2</i>	web-harvested	<i>web1+web2</i>	manual	218 (110/108)	50.5
<i>lab_htk_FOF</i>	laboratory-elicited	FOF interrogative and declarative	automatic	394 (193/201)	51
<i>lab_hand_FOF</i>	laboratory-elicited	FOF interrogative and declarative	manual	395 (194/201)	50.9
<i>lab_htk_declarative</i>	laboratory-elicited	FOF declarative	automatic	299 (144/155)	51.8
<i>lab_htk_interrogative</i>	laboratory-elicited	FOF interrogative	automatic	97 (49/48)	50.5
<i>lab_htk_SOF</i>	laboratory-elicited	SOF declarative	automatic	194 (94/101)	51.7

Table 1. Summary of datasets analyzed

Following convention in the machine learning community, I evaluated the classifiers by training them on one set of data and testing them on a new set of data. Table 2 summarizes how the datasets are used in training and testing, and indicates the section of the chapter which contains the experimental results.

<u>Training set</u>		<u>Test set</u>	<u>Section</u>
<i>web1</i>			(5.2)
	====>	<i>web2</i>	(5.2.2)
	====>	<i>lab_htk_FOF</i>	(5.2.3)
		<i>lab_hand_FOF</i>	
	====>	<i>lab_declarative</i>	(5.2.4)
		<i>lab_interrogative</i>	
	====>	<i>lab_SOF</i>	(5.2.5)
<i>lab_htk_FOF</i>			(5.3)
	====>	<i>web1+2</i>	(5.3.2)
	====>	<i>lab_htk_SOF</i>	(5.3.3)

Table 2. Summary of training set / test set pairs.

I used three measurements to quantify classifier performance. Generalization accuracy rate is the percentage of correctly classified tokens. This is compared against a baseline accuracy, which is a simple measure of the percentage of the most frequently occurring class (“s” or “ns”).

Thirdly, I calculated a balanced error rate, which takes into consideration the relative number of false positives and false negatives.

$$(11) \quad \text{Baseline accuracy} \quad \frac{\# \text{ tokens in largest class of test set}}{\# \text{ tokens in both classes in test set}}$$

$$(12) \quad \text{Generalization accuracy} \quad \frac{\# \text{ of tokens in test set accurately classified}}{\# \text{ of tokens in test set}}$$

$$(13) \quad \text{Balanced error rate}$$

$$\frac{\# \text{ incorrect "s"}}{\# \text{ total "s"}} + \frac{\# \text{ incorrect "ns"}}{\# \text{ total "ns"}} \times \frac{1}{2} \times 100$$

For a linear discriminant analysis (LDA) model, it is also possible to perform a Wilks' lambda test of significance. All the LDA models I present are significant ($p > 0.001$).

5.2 *Web-trained machine learning classifiers*

web1 ==>

5.2.1 *Feature selection*

Section 4.2 described the different methods of feature selection—automatic and manual—which I applied to the datasets¹⁶. I list the results of feature selection for the *web1* training set in Table 3.

¹⁶ In some cases, different iterations of the algorithm yielded different feature sets. The feature sets reported here reflect those yielding maximum classification accuracy.

Selection method	Number of features	Features selected
2. Best Duration (VarSelRF)	2	duration_V2, duration_syll1
3. Best Quantity (VarSelRF)	12	amp_ratio, energy_ratio, mean_f0_ratio, maxf0_ratio, maxf0Time_ratio, maxf0Time_V2, minf0_ratio, minf0Time_ratio, minf0Time_V2_percent, pulses_ratio, rangef0_ratio, rangef0_V2
4. Best Quality (VarSelRF)	2	f1f2Time40_V2, f1f2Time50_V2
5. 2-4 Combined	16	all of the above
6. Best Overall (VarSelRF)	4	duration_V2, mean_f0_ratio, f1f2Time40_V2, f1f2Time50_V2
7. Hand-picked A	3	duration_V2, duration_C3, f1f2Time50_V2
8. Hand-picked B	5	duration_V2, duration_C3, mean_f0_ratio, f1f2Time40_V2, f1f2Time50_V2
9. Hand-picked C	4	duration_V2, duration_C3, mean_f0_ratio, f1f2Time50_V2
10. Hand-picked D	3	duration_V2, duration_C3, f1f2Time40_V2
11. Best overall (F-score+SVM)	4	duration_V2, f1f2Time40_V2, f1f2Time50_V2, f1f2Time60_V2

Table 3. List of feature sets selected for training on dataset web1

Visualizing is usually the recommended first step in exploring quantitative data, and is often an excellent presentation aid. How to visualize machine learning classifiers, particularly classifiers that use many different features is a non-trivial matter. We cannot easily represent on a 2- or 3-dimensional graph of all the surfaces of the decision boundary. Graphical exploration may be instructive, however, in the form of scatter plot matrices, in which individual features are compared pairwise. A detailed pairwise graphical comparison of all 308 features is, of course, impractical. In this and subsequent subsections, I present a limited number of scatterplots in order to illustrate the predictive contribution of certain features.

Figure 5 is a scatterplot of the top duration measures for dataset *web1*. We can see that while all pairs offer some separation of the data into the two classes, ‘duration_V2’ and ‘duration_C3’ offer the best pairwise separation. Tokens of class “s” show a strong tendency towards greater duration of *I*.

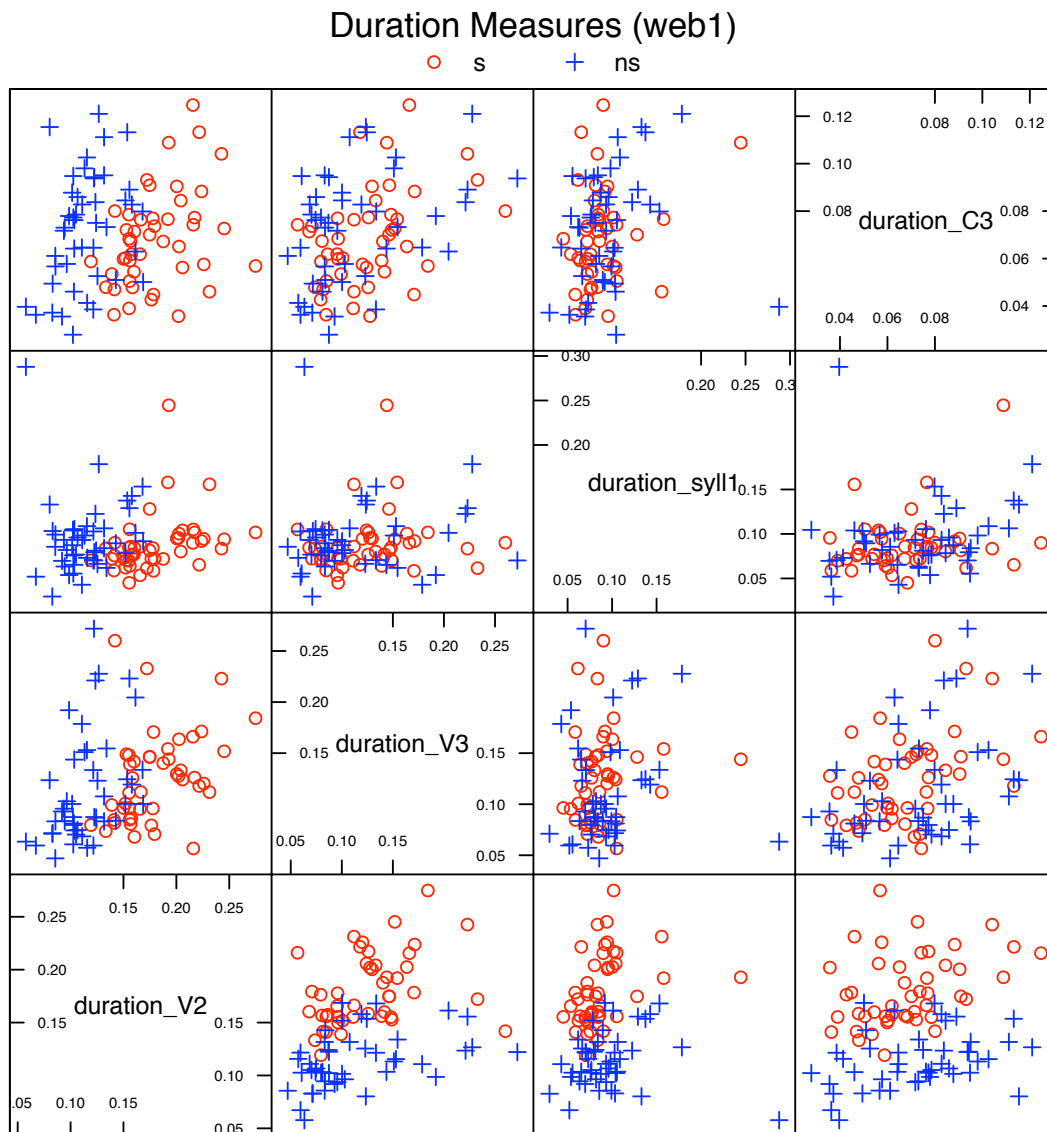


Figure 5. Scatterplot matrix of four duration-related measures for *web1*, with 1 outlier removed. From left to right: duration (ms) of *I*, duration (ms) of vowel in *did*, duration (ms) of vowel and nasal in *than*, duration (ms) of second stop consonant in *did*.

In a pair-wise comparison of four of the top quantity-related measures, the data are considerably denser. Because the measures are ratios (i.e. a value measured from *I* divided by a value measured from *did*), plotting their log values helps to reveal some limited separation between classes (Figure 6). We may observe general trends, such as a higher maximum *f0* ratio,

f_0 range ratio, energy ratio and amplitude ratio for tokens of class “s”; however, there is considerable overlap between classes and no clear boundary is evident.

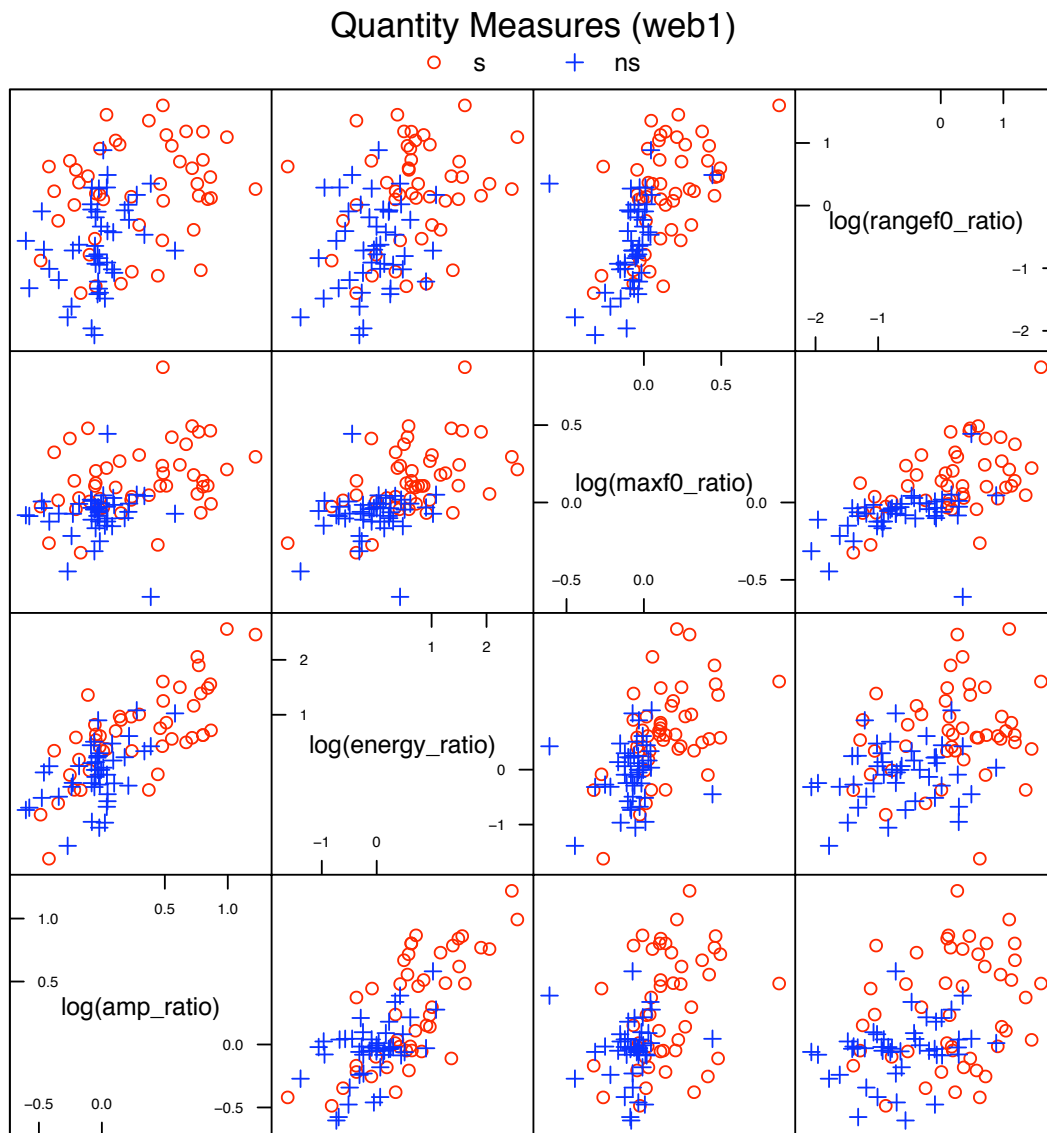


Figure 6. Scatterplot matrix of four quantity-related measures for *web1*. From left to right: log value of the ratio of mean amplitudes in vowels of *I* and *did*, log of ratio of mean energy over vowels of *I* and *did*, log of ratio of f_0 maxima in vowels of *I* and *did*, log of ratio of f_0 range (i.e. maximum - minimum) over vowels of *I* and *did*.

Pair-wise comparison of quality-related measures—the differences in first and second formant values of V2—show good separation of classes (Figure 7). Tokens in class “s” tend to have a small formant differential; tokens in class “ns” have a large formant differential. A small

F2-F1 formant differential is consistent with a more canonical or hyperarticulated [a] phone, the first part of the [aI] diphthong of *I*. Note that the plotted values fall along a line, since formant values at, e.g. 40 and 50 percent of the vowel, are proportionally related: a token with a large F2-F1 differential at 40 percent tends to have a similarly large F2-F1 differential at 50 percent. The relationship is of course less robust between the F2-F1 differential at 30 and 60 percent.

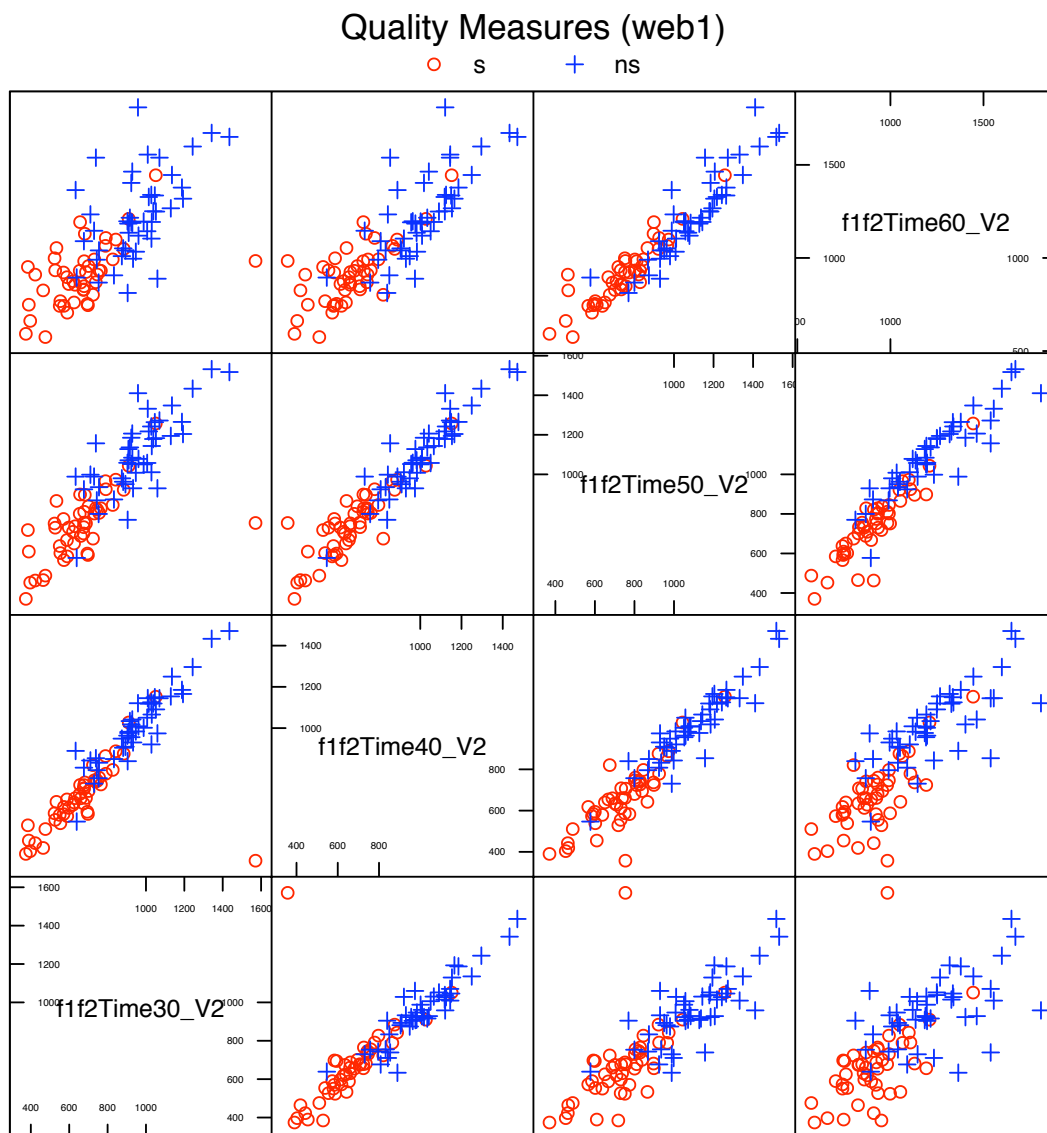


Figure 7. Scatterplot matrix of four quality-related measures for *web1*. From left to right: formant differentials (F2-F1) (Hz) at 30%, 40%, 50% and 60% into the duration of *I*.

The next two scatterplots (Figures 8 and 9) compare best algorithm- and experimenter-selected features, respectively. One can see that ‘duration_V2’ plotted against ‘f1f2Time40_V2’ or ‘f1f2Time50_V2’ demonstrates excellent separation, as does ‘duration_V2’ against ‘duration_C3’.

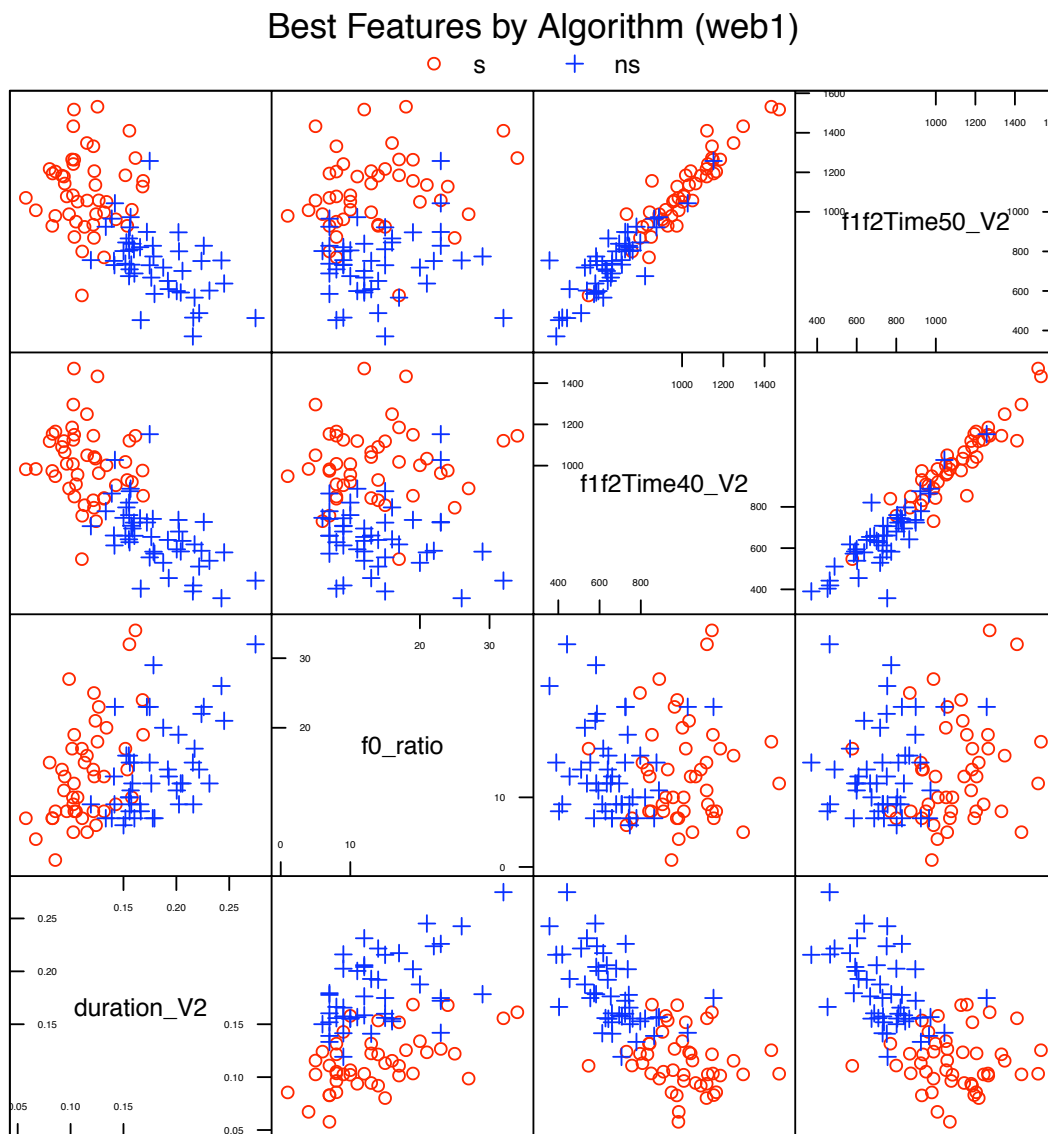


Figure 8. Scatterplot matrix of four best measures chosen automatically using the VarSelRF algorithm for *web1*. From left to right: duration of *I*, ratio of mean *f0* over vowels in *I* and *did*, formant differentials (F2-F1) at 40% into the duration of *I*, formant differentials (F2-F1) at 50% into the duration of *I*.

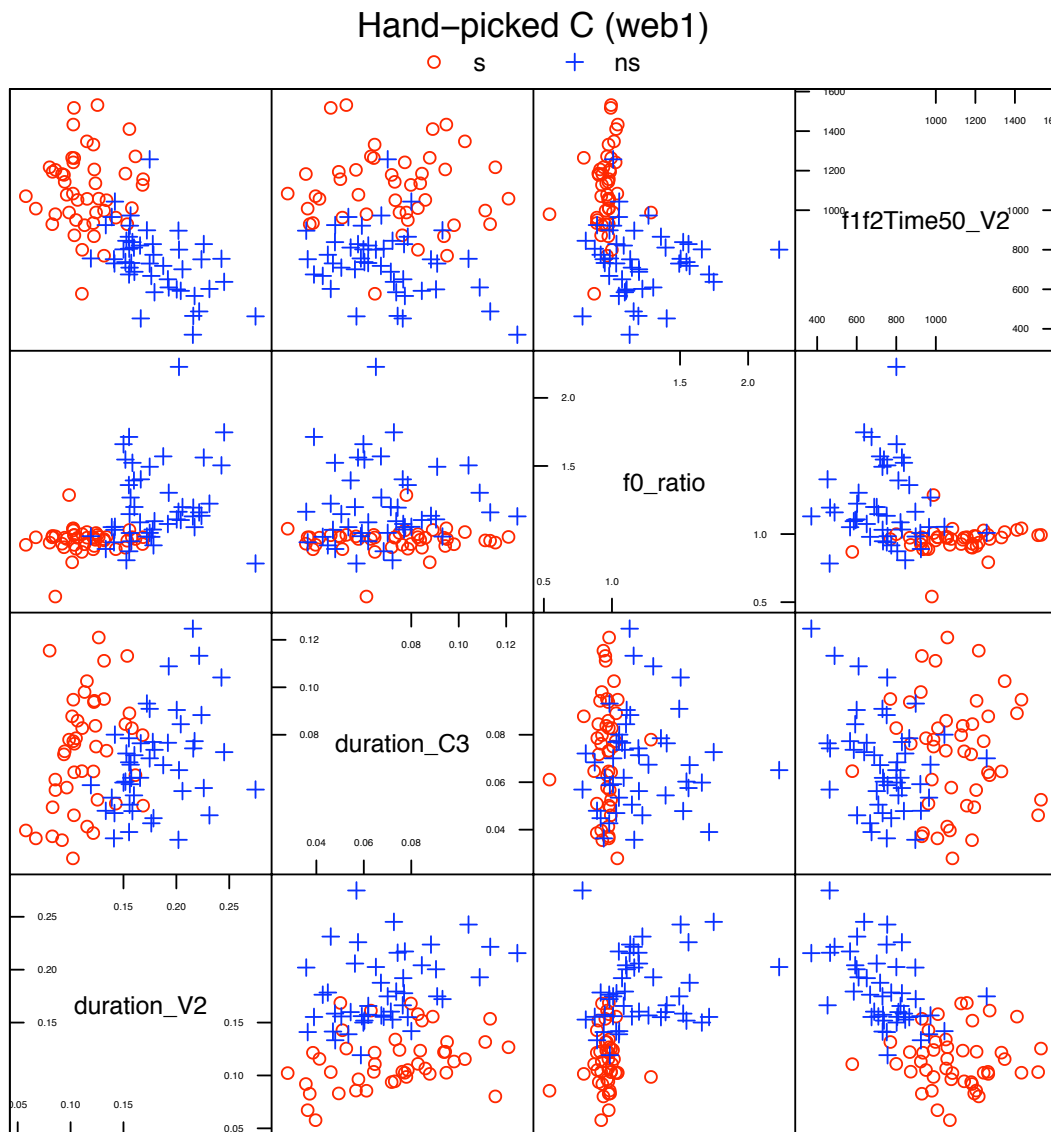


Figure 9. Scatterplot matrix of four experimenter-selected measures ('Hand-picked C') for *web1*. From left to right: duration of *I*, duration of first stop closure in *did*, ratio of mean *f0* in *I* and *did*, formant differentials (F2-F1) at 50% into the duration of *I*,

Examination of these scatterplots also suggests that while some measures allow better separation of data than others, there is no single measurement or measurement pair that separates all of the data reliably. A sophisticated representation of the data using several acoustic measures is necessary for optimal classification.

5.2.2 Classification accuracy on web-harvested data

web1 ==> *web2*

In this section we train the classifiers on a web-harvested dataset, *web1*, and test the classifiers on a second web-harvested dataset, *web2*. All of these web-trained models achieved generalization accuracies and balanced error rates well above the baseline (baseline accuracy 51.2). The different machine learning methods and feature sets yielded classifiers which are quite competitive. Among the different feature sets, classifiers using quantity features performed the weakest and, in general, the classifiers which used experimenter-selected features outperformed both the classifiers which used the complete set of features and the classifiers which used automatically-selected feature sets. The best performing classifier, at 92.9% generalization accuracy and 6.5% BER, was the radial SVM trained with experimenter-selected feature set C: ‘duration_V2’, ‘duration_C3’, ‘flf2Time50_V2’ and ‘mean_f0_ratio’.

Feature set	web2 (corpus)			
	Baseline	SVM (RBF)	SVM (linear)	LDA
1. Full	51.2	82.7 (14.3)	85.0 (14.2)	--
2. Best Duration (VarSelRF)	51.2	80.3 (17.8)	85.8 (12.7)	85.0 (14.2)
3. Best Quantity (VarSelRF)	51.2	82.7 (17.0)	80.3 (19.6)	78.7 (21.1)
4. Best Quality (VarSelRF)	51.2	78.7 (21.3)	79.5 (20.5)	78.7 (21.2)
5. 2-4 Combined	51.2	89.8 (10.1)	89.0 (10.3)	90.6 (9.0)
6. Best of full (VarSelRF)	51.2	86.6 (13.1)	90.6 (9.2)	87.4 (11.9)
7. Hand-picked A	51.2	91.3 (7.7)	92.1 (7.1)	91.3 (7.7)
8. Hand-picked B	51.2	92.1 (7.1)	92.1 (7.1)	91.3 (7.7)
9. Hand-picked C	51.2	92.9 (6.5)	92.1 (7.1)	91.3 (7.7)
10. Hand-picked D	51.2	89.0 (9.9)	91.3 (7.7)	90.6 (8.3)
11. Best of full (F-score+SVM)	51.2	87.4 (12.5)	86.6 (13.3)	85.8 (14.1)

Table 4. Accuracy rate (and balanced error rate) for different models: training set web1; test set web2.

5.2.3 Classification accuracy on laboratory-elicited data

web1 ==> *lab_htk_FOF*

lab_hand_FOF

Next, we use the same set of web-trained classifiers and test them on laboratory data, laboratory data annotated automatically (*lab_htk_FOF*) and laboratory data annotated by hand (*lab_hand_FOF*).

First, let's consider laboratory data annotated automatically (Table 4). Generally, the web-trained classifiers performed less well on laboratory-elicited data, although all of the web-trained classifiers performed well above a baseline of 51.0%. Again, the models based on different classification methods and feature sets remained competitive with each other. The best-performing classifier, at 87.3% generalization accuracy and 11.0% BER, was a linear SVM using experimenter-selected feature set A: 'duration_V2', 'duration_C3' and 'flf2Time50_V2'. In general, the web-trained classifiers performed better on web data than on these laboratory data.

Feature set	Baseline	lab_htk_FOF		
		SVM (RBF)	SVM (linear)	LDA
1. Full	51.0	72.3 (19.9)	82.2 (16.3)	--
2. Best Duration (VarSelRF)	51.0	84.5 (13.1)	85.8 (12.1)	84.0 (12.6)
3. Best Quantity (VarSelRF)	51.0	72.3 (25.3)	60.4 (39.6)	53.0 (46.9)
4. Best Quality (VarSelRF)	51.0	73.9 (20.1)	75.6 (18.8)	76.9 (18.7)
5. 2-4 Combined	51.0	81.5 (18.1)	84.5 (14.4)	74.4 (21.8)
6. Best of full (VarSelRF)	51.0	83.8 (16.1)	74.1 (20.9)	71.3 (21.9)
7. Hand-picked A	51.0	85.8 (11.9)	87.3 (11.0)	84.3 (12.9)
8. Hand-picked B	51.0	79.4 (17.3)	75.4 (20.4)	74.9 (19.6)
9. Hand-picked C	51.0	82.2 (16.8)	74.6 (21.2)	75.1 (19.4)
10. Hand-picked D	51.0	82.2 (16.6)	75.1 (20.6)	77.2 (17.6)
11. Best of full (F-score+SVM)	51.0	79.2 (15.2)	83.5 (13.4)	81.0 (14.9)

Table 5. Accuracy rate (and balanced error rate) for different models: training set *web1*; test set *lab_htk_FOF*.

Now let's consider the laboratory data annotated by hand (Table 6). In general, the classifier performances did *not* improve for this hand-annotated test set. Compare the linear SVM classifier using experimenter-selected feature set A, which achieved 83.5% accuracy and 13.5% BER on the hand-annotated data. The corresponding classifier was the best-performing classifier on the automatically annotated laboratory data (Table 5) at 87.3% generalization accuracy and 11.0% BER.

		lab_hand_FOF		
Feature set	Baseline	SVM (RBF)	SVM (linear)	LDA
1. Full	51.0	75.9 (80.8)	80.8 (14.4)	--
2. Best Duration (VarSelRF)	51.0	79.0 (16.6)	79.7 (16.6)	79.2 (16.4)
3. Best Quantity (VarSelRF)	51.0	73.9 (24.5)	62.3 (37.7)	53.4 (46.5)
4. Best Quality (VarSelRF)	51.0	72.2 (21.1)	73.7 (73.7)	74.9 (19.8)
5. 2-4 Combined	51.0	81.5 (18.3)	83.5 (14.4)	73.4 (22.2)
6. Best of full (VarSelRF)	51.0	82.5 (17.3)	72.9 (21.0)	71.6 (20.6)
7. Hand-picked A	51.0	79.5 (15.4)	82.8 (13.8)	79.2 (15.5)
8. Hand-picked B	51.0	76.5 (18.6)	74.2 (19.9)	73.0 (19.3)
9. Hand-picked C	51.0	79.7 (18.1)	72.9 (21.0)	72.9 (19.9)
10. Hand-picked D	51.0	82.0 (14.2)	83.5 (13.5)	80.5 (14.8)
11. Best of full(F-score+SVM)	51.0	75.9 (17.2)	80.8 (14.4)	78.0 (16.5)

Table 6. Accuracy rate (and balanced error rate) for different models: training set web1; test set lab_hand_FOF.

We may ask why the accuracy on the lab data is not more similar to the accuracy on the web data. Scatterplots of the laboratory data are instructive. The data are more densely distributed than the web data, in part due to the number of tokens (394 in *lab_htk_FOF* compared with 91 in *web1*). However, it is clear from review of these scatterplots that the various phonetic features do not allow separation of classes in pairwise comparison to the same extent as the web corpus data. In the remainder of this subsection I recreate, for the laboratory data, the scatterplots of the web data we reviewed in Section 5.2.1.

Beginning again with the features of duration, note that the annotation by HTK results in duration values rounded to the nearest millisecond. This rounding is a feature of the forced-alignment procedure, and is the cause of the “binned” distribution of duration in Figure 10. Compare this distribution with the distribution of duration values from the hand-annotated lab data in Figure 11. In spite of the high density distribution observable in both datasets, the classifiers using duration values alone (Model 2 – ‘duration_V2’, ‘duration_syll1’, ‘duration_V3’, ‘duration_C3’) performed well, particularly on the automatically annotated data.

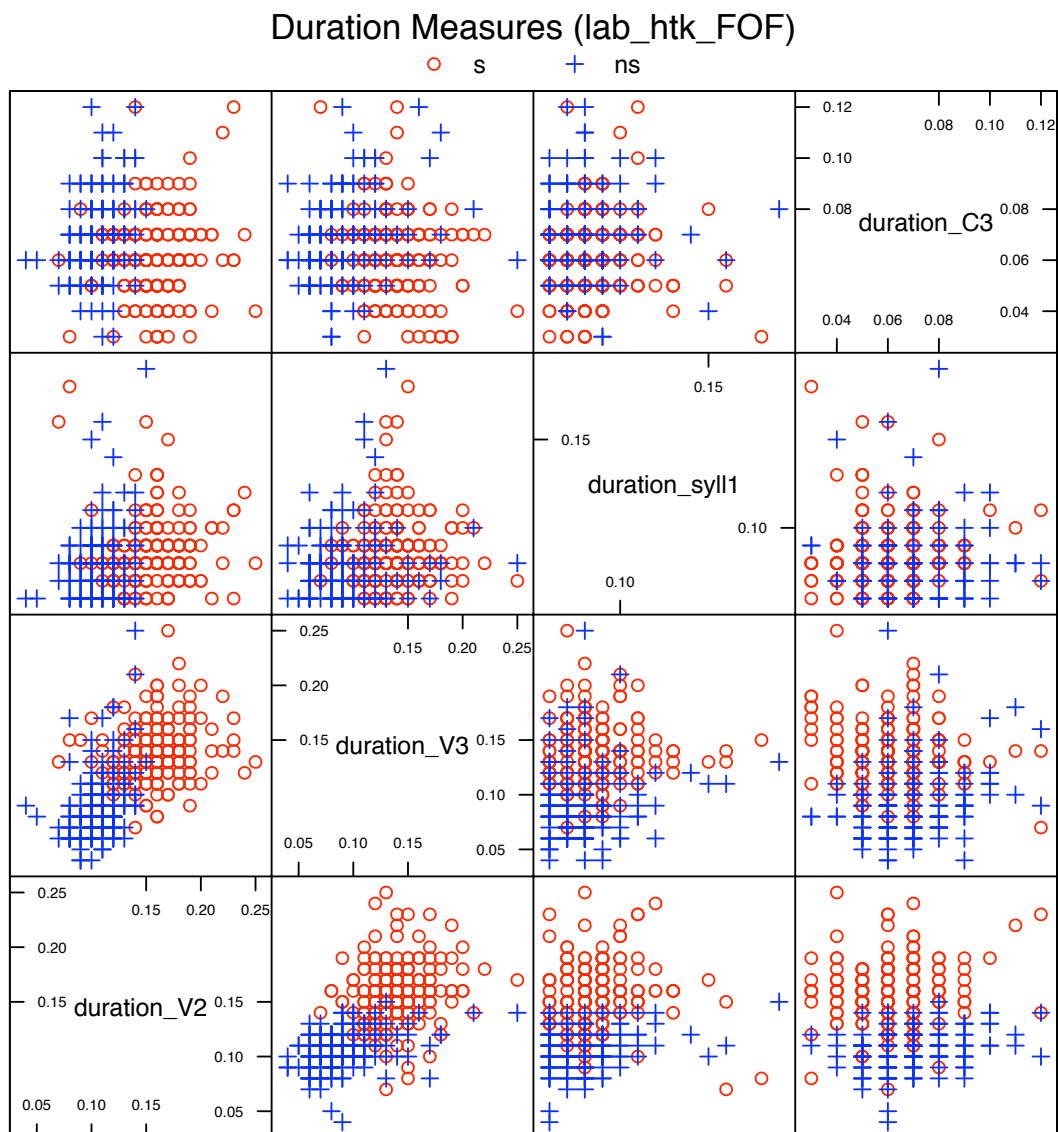


Figure 10. Scatterplot matrix of four duration measures for *lab_htk_FOF*. From left to right: duration of *I* (ms), duration of vowel in *did*, duration of vowel and nasal in *than*, duration of first stop closure in *did*.

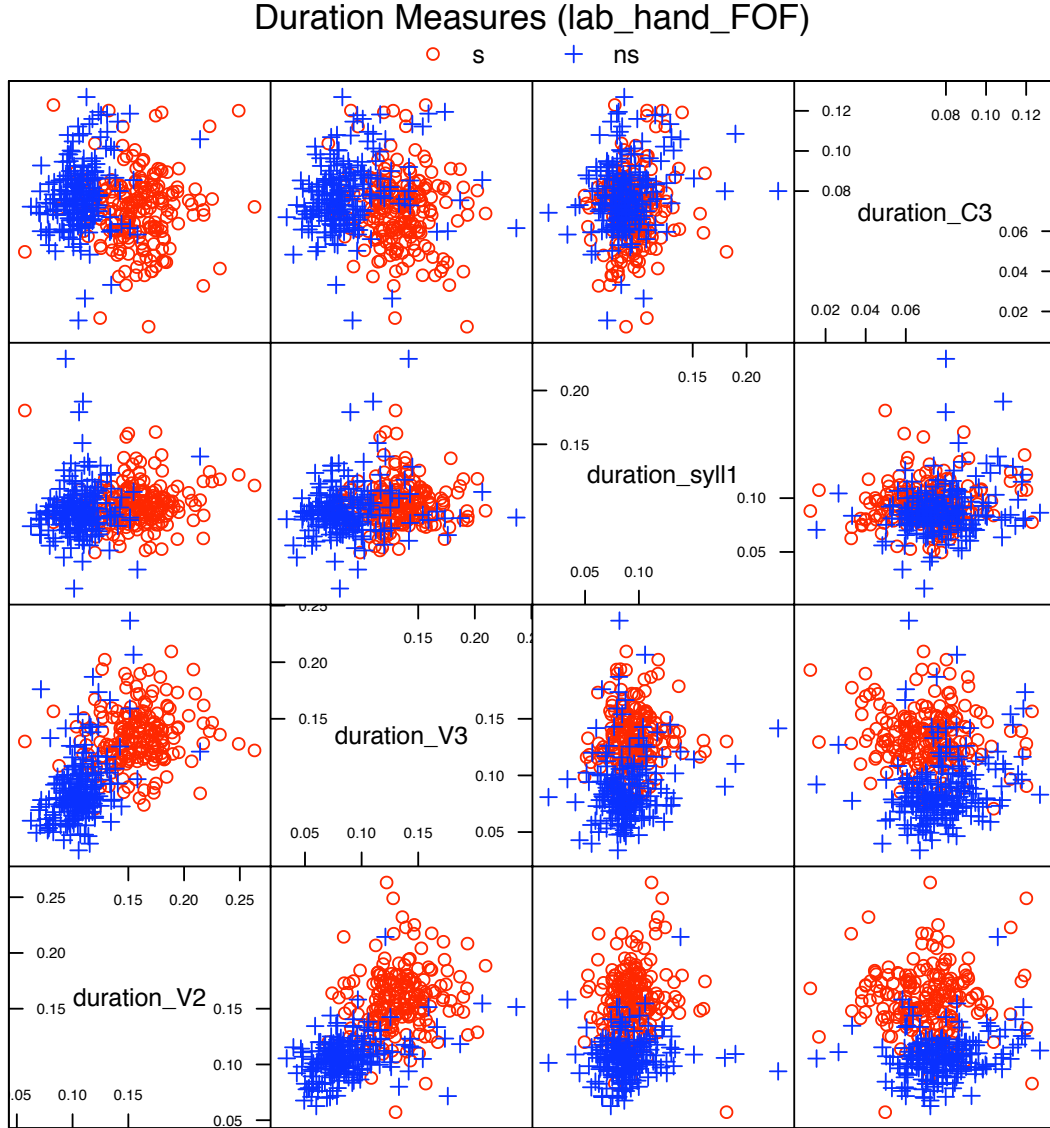


Figure 11. Scatterplot matrix of four duration measures for *lab_hand_FOF* (1 outlier removed). From left to right: duration of *I*, duration of vowel in *did*, duration of vowel and nasal in *than*, duration of first stop closure in *did*.

The automatic extraction of *f0* information is notoriously difficult, and errors in annotation have the potential to yield even more erroneous or undefined *f0* measures. Indeed, the hand-annotated data contained far fewer errors in extraction of *f0* measures. For example, 16.5% of *lab_htk_FOF* tokens had undefined values for the attribute *mean_f0_ratio* (the ratio of *f0* means in V2 and V3). Of those, the declarative tokens accounted for 11.9% and the question tokens

4.6% By comparison, a mere 1.8% of tokens in the hand-annotated dataset *labhand_FOF* had undefined values for *mean_f0_ratio*, comparable with 1.6% in *web2*, which was also hand-annotated.

Models of quantity-related measures performed equally poorly on both datasets, the LDA models achieving little better than chance accuracy at 53.0% (*lab_htk_FOF*) and 53.4% (*lab_hand_FOF*). Scatterplots of densely distributed data corroborate the non-predictiveness of these measures for the laboratory datasets (Figures 12 and 13). For example, class “s” shows a tendency for greater energy ratio and greater amplitude ratio, but there is considerable overlap.

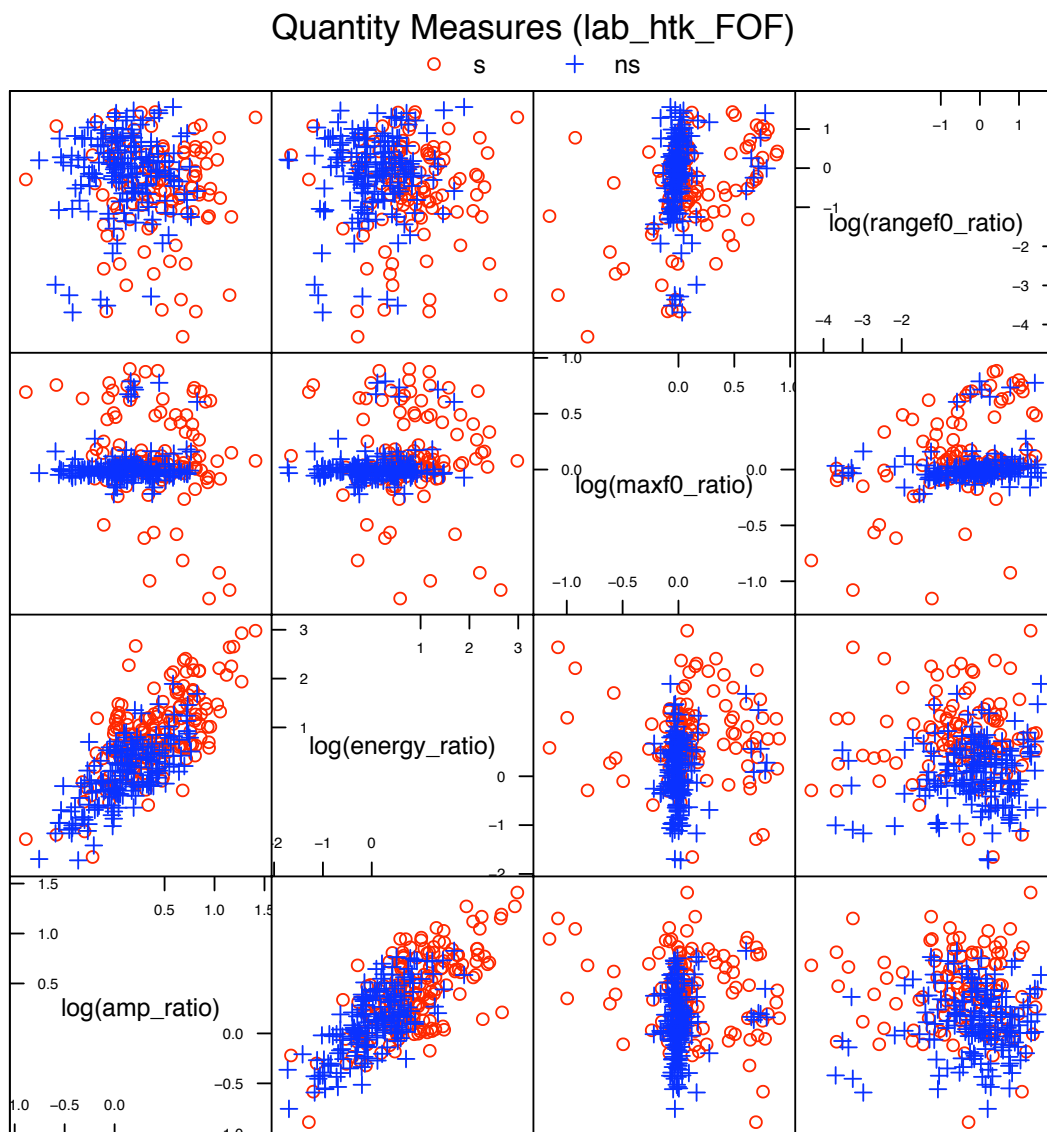


Figure 12. Scatterplot matrix for four quantity-related measures for *lab_htk_FOF* (5 outliers removed). From left to right: log value of the ratio of mean amplitudes in vowels of *I* and *did*, log of ratio of mean energy over vowels of *I* and *did*, log of ratio of *f0* maxima in vowels of *I* and *did*, log of ratio of *f0* range (i.e. maximum - minimum) over vowels of *I* and *did*.

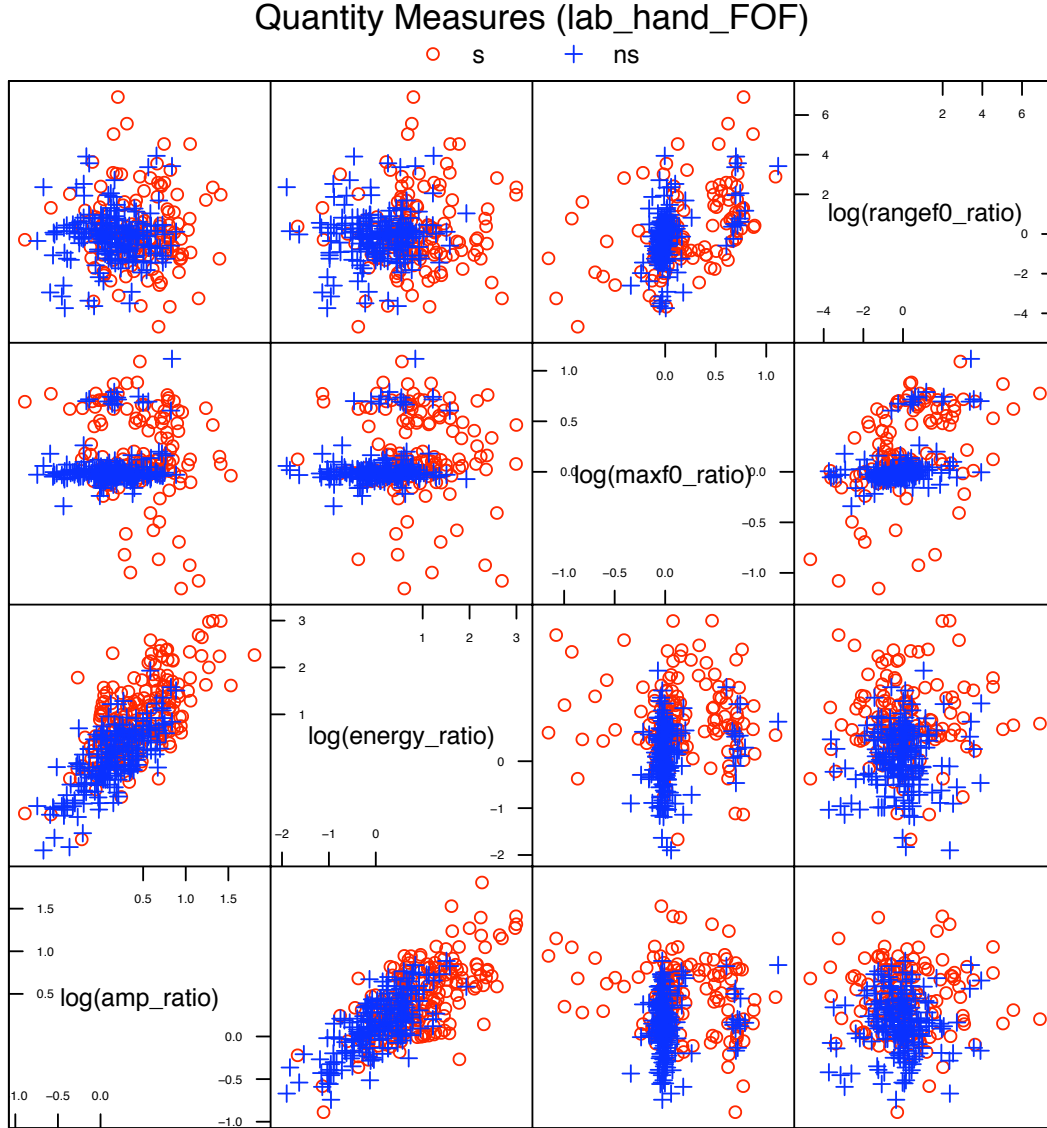


Figure 13. Scatterplot matrix for four quantity-related measures for *lab_hand_FOF*. From left to right: log value of the ratio of mean amplitudes in vowels of *I* and *did*, log of ratio of mean energy over vowels of *I* and *did*, log of ratio of *f0* maxima in vowels of *I* and *did*, log of ratio of *f0* range (i.e. maximum - minimum) over vowels of *I* and *did*.

The classifiers using quality-related measures alone performed well above the baseline (from 72.2% to 76.9% generalization accuracies), although not with the same level of accuracy as the classifiers using durational measures alone. The formant differentials plotted in Figures 14 and 15 show an overlap of classes, they also show a clear tendency towards separation.

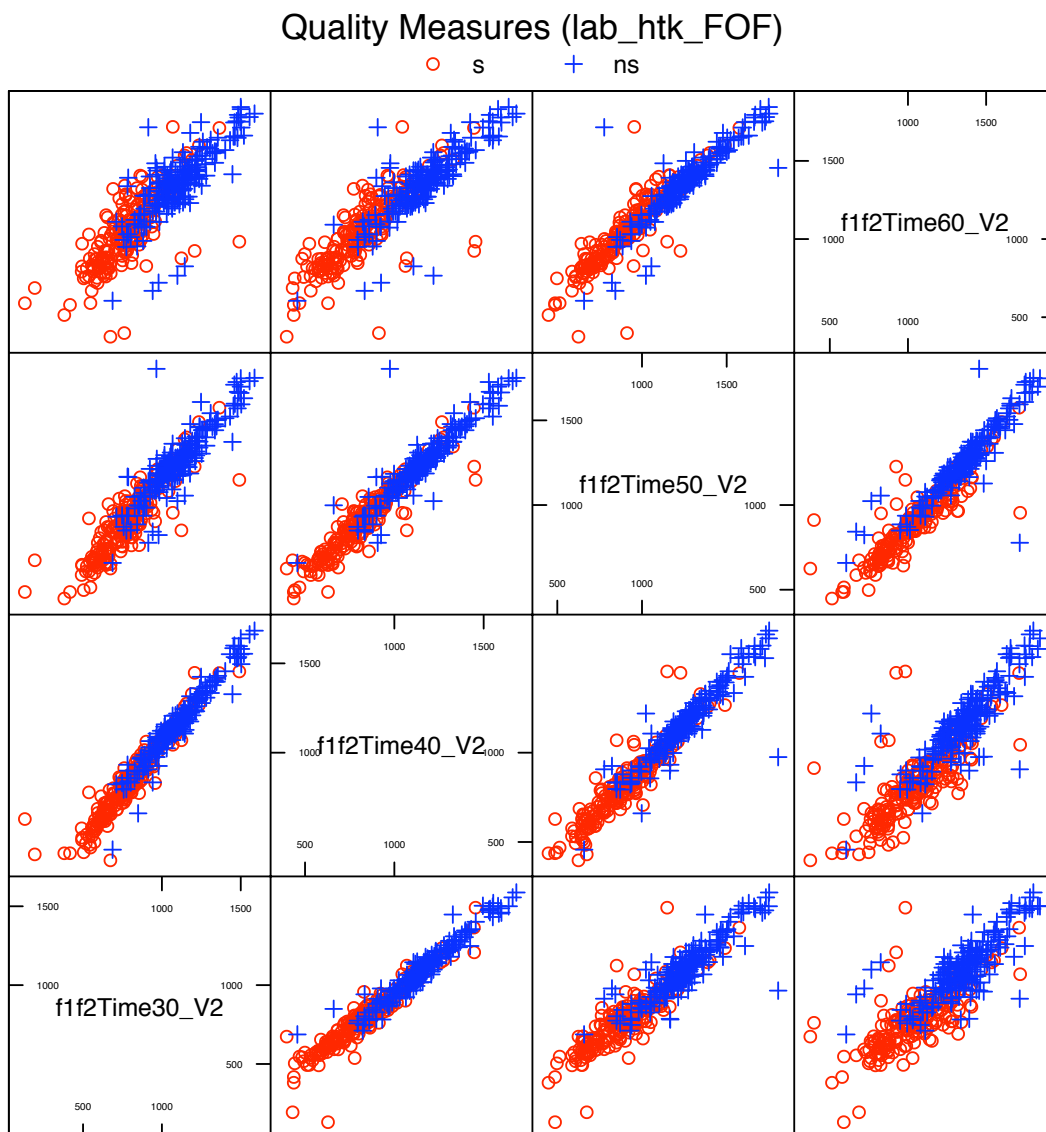


Figure 14. Scatterplot matrix of four quality-related measures for *lab_htk_FOF*. From left to right: formant differentials (F2-F1) (Hz) at 30%, 40%, 50% and 60% into the duration of I.

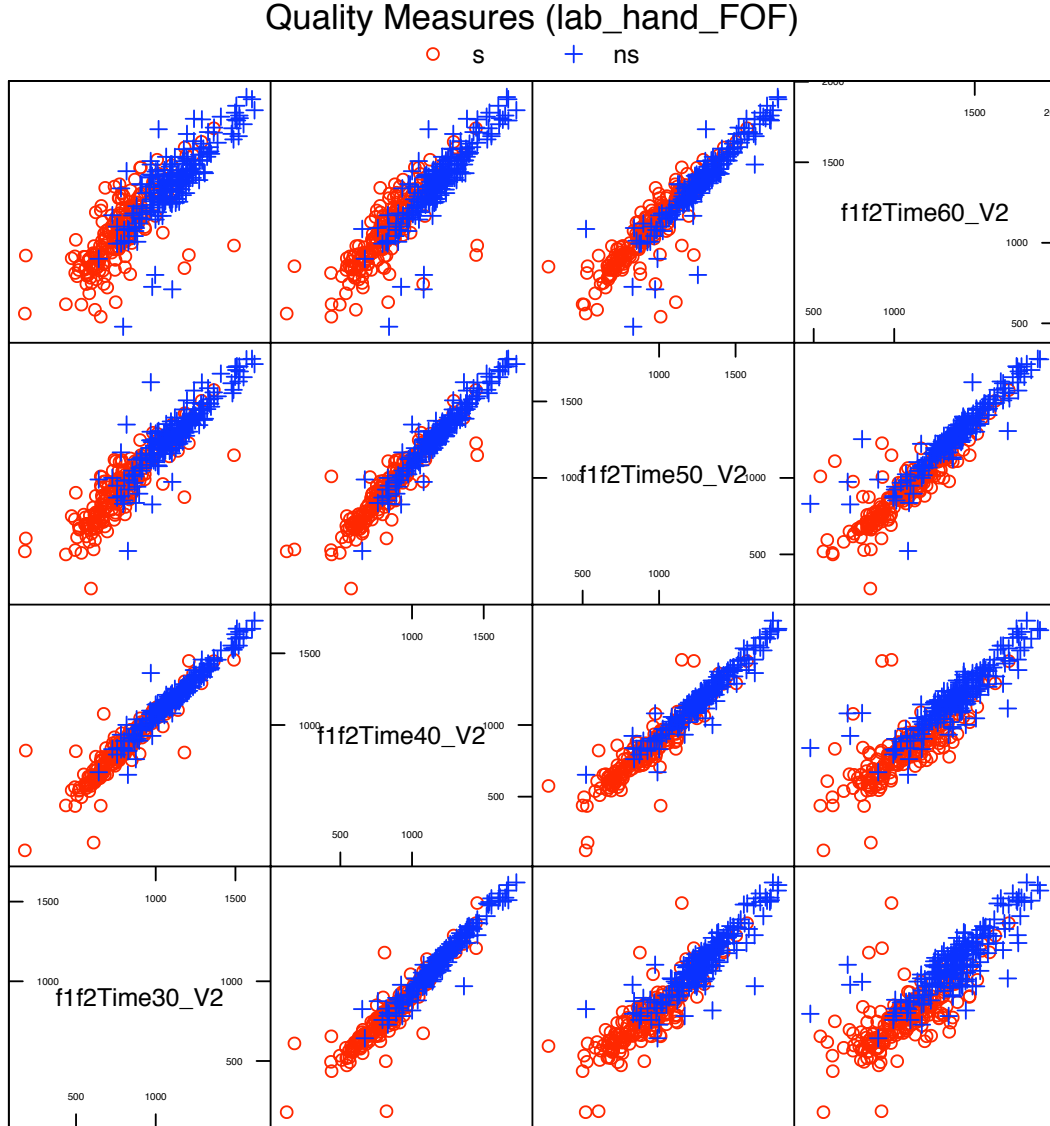


Figure 15. Scatterplot matrix of four quality-related measures for *lab_hand_FOF*. From left to right: formant differentials (F2-F1) (Hz) at 30%, 40%, 50% and 60% into the duration of I.

Finally, let us visualize distributions of the laboratory data for the features used in the best performing models. The best-performing classifier on laboratory data used experimenter-selected feature set Hand-picked D (‘duration_V2’, ‘duration_C3’, ‘f1f2Time40_V2’). Note that no *f0*-related measure is included in this set. For the sake of comparison with the web-harvested data, however, I’ll plot the features in Hand-Picked C (‘duration_V2’, ‘duration_C3’, ‘f0_ratio’, ‘f1f2Time50’). This set was used by the best-performing classifier on the web data, and was also

used by a high-performing classifier on the laboratory data. It is clear from figures 15 and 16 that these variables do not allow as neat a separation of classes for the laboratory data. The line that occurs in the scatterplots with ‘f0_ratio’ are due to tokens undefined for this value which have been coded as zero (cf. Section 4.1).

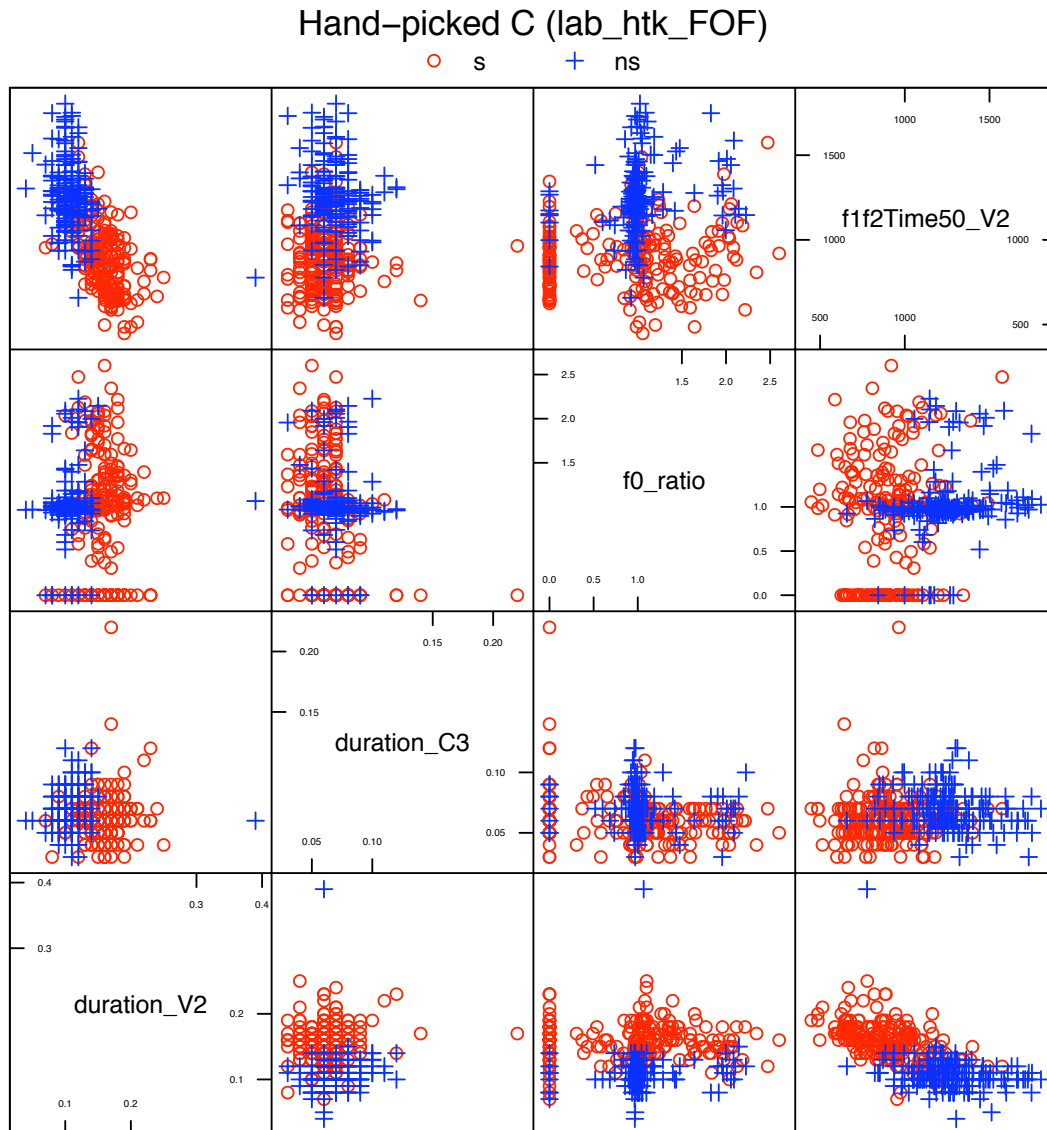


Figure 16. Scatterplot matrix of four experimenter-selected measures (‘Hand-picked C’) for *lab_htk_FOF*. From left to right: duration of *I*, duration of first stop closure in *did*, ratio of mean *f0* in *I* and *did*, formant differentials (F2-F1) at 50% into the duration of *I*,

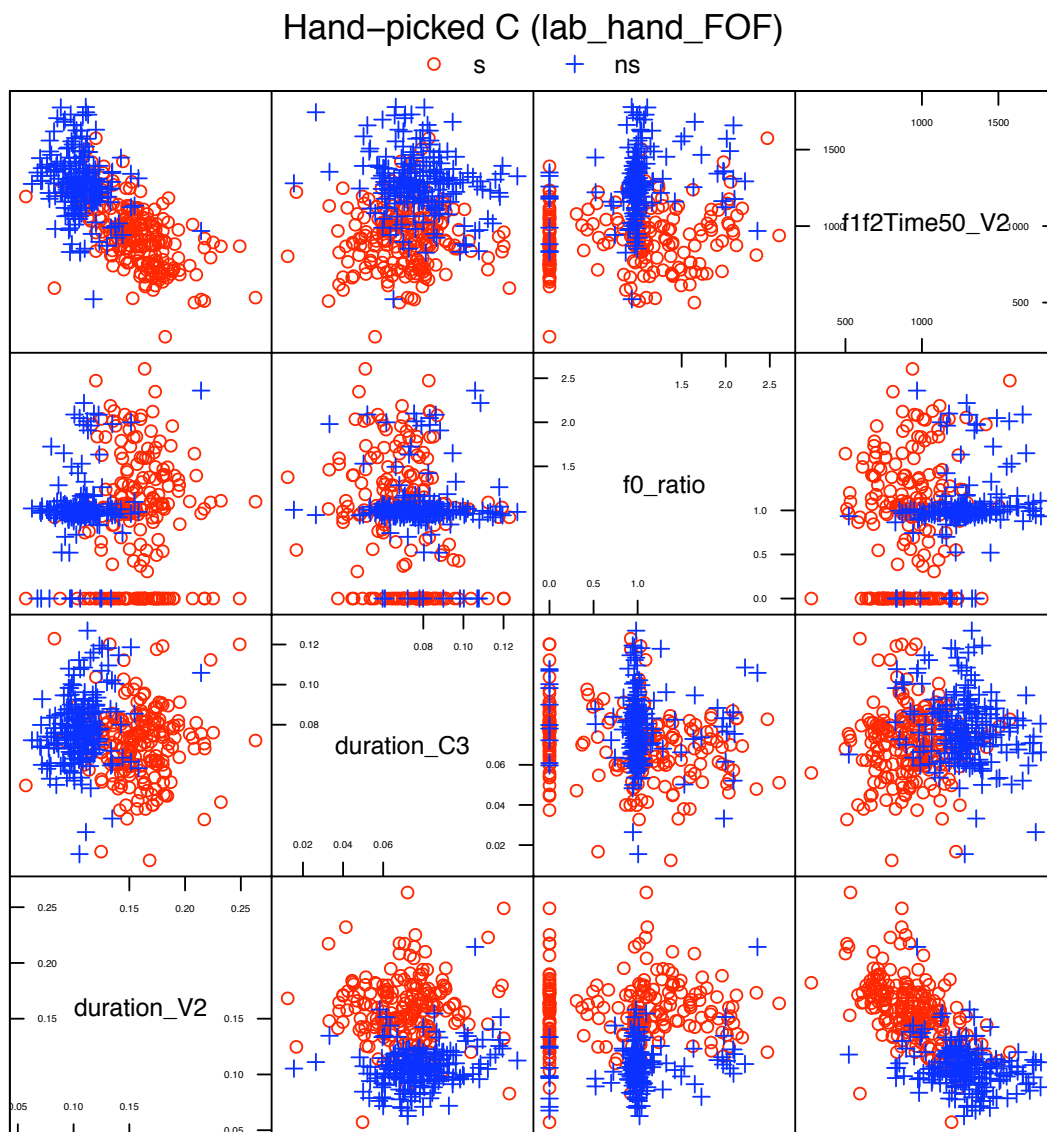


Figure 17. Scatterplot matrix of four experimenter-selected measures (‘Hand-picked C’) for *lab_hand_FOF*. From left to right: duration of *I*, duration of first stop closure in *did*, ratio of mean *f0* in *I* and *did*, formant differentials (F2-F1) at 50% into the duration of *I*.

In summary, the web-trained classifiers performed competitively on the laboratory data, but they performed even better on the web data. The hand-annotated laboratory data did not improve classifier performance over the automatically annotated laboratory data. And the classifiers which performed best on laboratory data were those lacking *f0* features: Hand-picked

A ('duration_V2', 'duration_C3', 'f1f2Time50_V2') and Hand-picked D ('duration_V2', 'duration_C3', 'f1f2Time40_V2').

5.2.4 Declaratives and interrogatives

web1 ==> thanIdidlab_declaratives

The laboratory production data we considered in the previous section are further divided into two experimental conditions: declaratives (299 tokens) and interrogatives (97 tokens). Since the hand-annotated data did not improve classifier performance, let's consider only the automatically aligned data

Focus in question contexts is typically realized with a low pitch accent (cf. 3-4). Therefore, the f_0 minimum is thought to be more phonetically salient than an f_0 maximum. Figure 17 shows an example of an f_0 contour over an utterance of “than I did” with subject focus in an interrogative context. The f_0 maximum occurs at the right edge of the vowel in *I* and is not predicted to be linguistically relevant. Figure 18 shows an example of an f_0 contour over an utterance of “than I did” with subject focus in a declarative context. Similarly, the f_0 minimum occurs at the right edge of the vowel in *I* and is not predicted to be linguistically relevant.

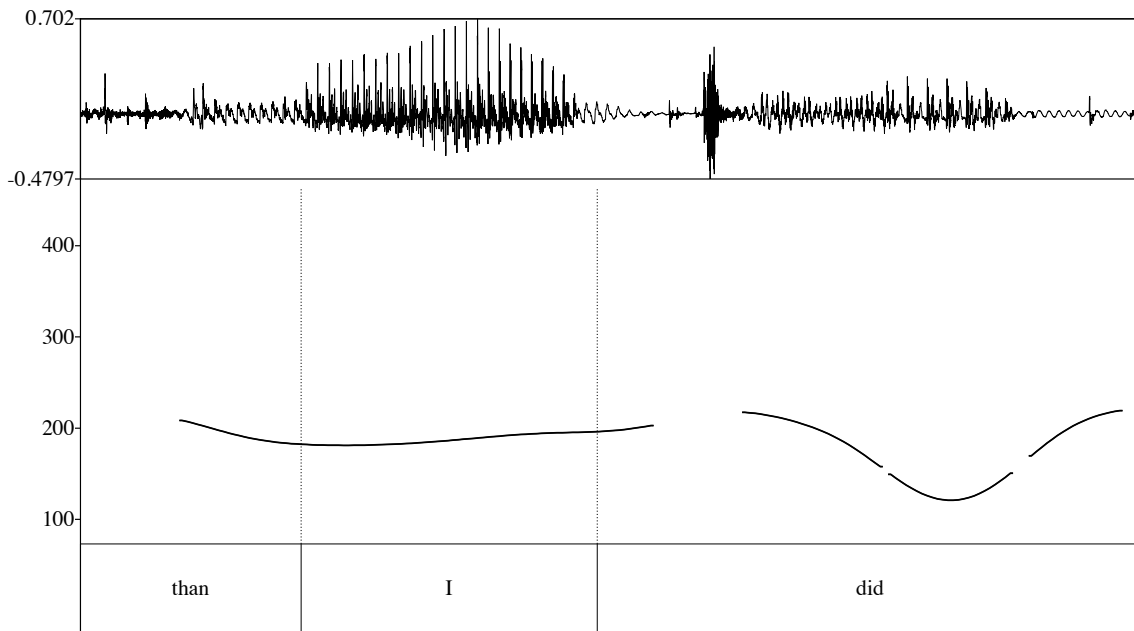


Figure 18. Smoothed f_0 contour and waveform from an elicited utterance of "than I did" (subject focus) in an interrogative context.

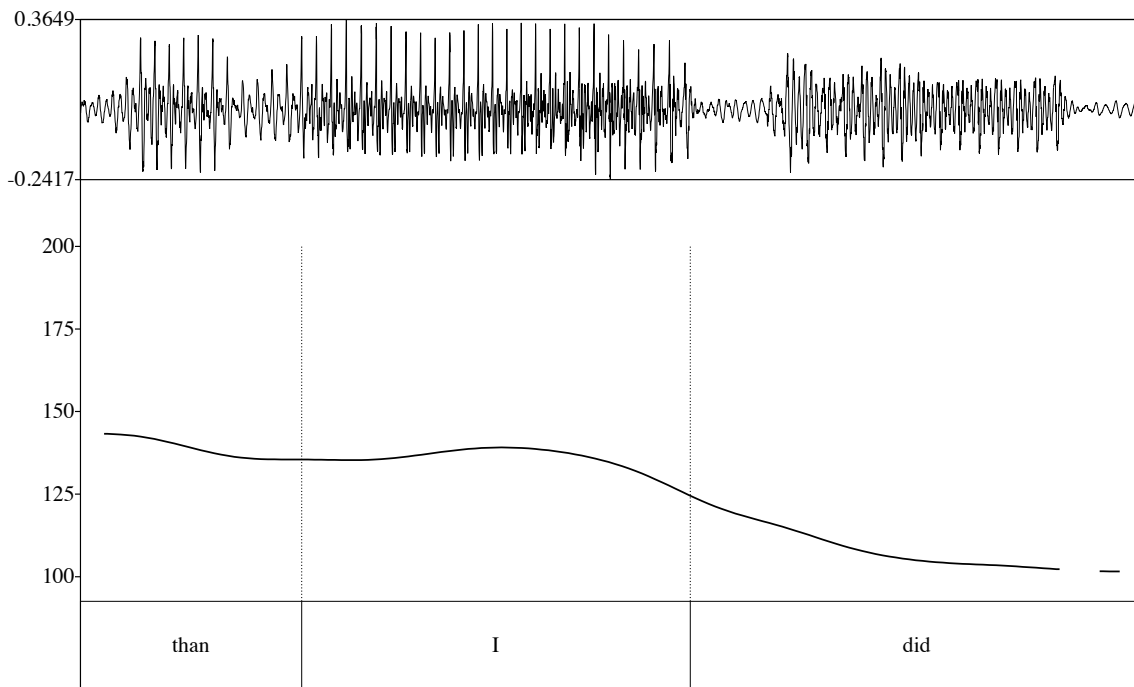


Figure 19. Smoothed f_0 contour and waveform from a naturally occurring utterance of "than I did" (subject focus) in a declarative context.

Since the classifiers are trained on data from declarative contexts exclusively, we might expect the classifiers, particularly those using quantity-related measures (including $f0$) exclusively, to perform better on data from declarative contexts than data from question contexts. Indeed, the classifiers using quantity-related features performed poorly on data from both experimental conditions: as low as 50.2% and no higher than 72.7% generalization accuracy. Moreover, the best performing classifiers were those which lacked $f0$ features altogether: Hand-picked A (‘duration_V2’, ‘duration_C3’, ‘f1f2Time50_V2’).

These results are consistent with the results for *lab_htk_FOF*, for which the experimental conditions declarative and interrogative are pooled. In other words, a classifier is able to successfully classify across both contexts by using features other than those related to $f0$.

Classifier performance on the declarative and interrogative datasets is summarized in Tables 7 and 8.

Feature set	lab_htk_declarative			
	Baseline	SVM (RBF)	SVM (linear)	LDA
1. Full	51.5	72.1 (20.6)	80.1 (18.2)	--
2. Best Duration (VarSelRF)	51.5	83.5 (13.7)	86.2 (11.6)	84.2 (12.6)
3. Best Quantity (VarSelRF)	51.5	72.7 (26.0)	60.3 (39.7)	50.2 (49.7)
4. Best Quality (VarSelRF)	51.5	74.4 (19.8)	75.8 (18.6)	76.4 (19)
5. 2-4 Combined	51.5	80.5 (19.4)	84.2 (14.6)	71.0 (24.6)
6. Best of full (VarSelRF)	51.5	82.5 (17.2)	71.7 (22.6)	69.7 (22.9)
7. Hand-picked A	51.5	85.2 (12.3)	86.9 (11.4)	83.8 (13.2)
8. Hand-picked B	51.5	77.4 (18.7)	73.4 (21.6)	72.1 (21.6)
9. Hand-picked C	51.5	80.5 (18.2)	72.4 (22.6)	72.4 (21.4)
10. Hand-picked D	51.5	80.8 (17.7)	73.1 (21.8)	74.7 (19.2)
11. Best of full (F-score+SVM)	51.5	79.5 (14.9)	83.5 (13.4)	81.5 (14.6)

Table 7. Accuracy rate (and balanced error rate) for different models: training set web1; test set lab_htk_declarative.

		lab_htk_interrogative		
Feature set	Baseline	SVM (RBF)	SVM (linear)	LDA
1. Full	50.5	73.2 (17.3)	88.7 (10.5)	--
2. Best Duration (VarSelRF)	50.5	87.6 (11.3)	84.5 (13.5)	83.5 (12.3)
3. Best Quantity (VarSelRF)	50.5	71.1 (20.4)	60.8 (38.4)	61.9 (37.4)
4. Best Quality (VarSelRF)	50.5	72.2 (21.4)	75.3 (19.5)	78.4 (17.6)
5. 2-4 Combined	50.5	84.5 (13.6)	85.6 (13.9)	84.5 (13.5)
6. Best of full (VarSelRF)	50.5	87.6 (12.3)	81.4 (15.6)	76.3 (18.9)
7. Hand-picked A	50.5	87.6 (10.6)	88.7 (9.9)	85.6 (12.0)
8. Hand-picked B	50.5	85.6 (12.8)	81.4 (16.5)	83.5 (13.4)
9. Hand-picked C	50.5	87.6 (12.1)	81.4 (16.5)	83.5 (13.4)
10. Hand-picked D	50.5	86.6 (13)	81.4 (16.5)	84.5 (12.7)
11. Best of full (F-score+SVM)	50.5	78.4 (16.4)	83.5 (13.4)	79.4 (15.8)

Table 8. Accuracy rate (and balanced error rate) for different models: training set web1; test set thanIdidlab_interrogative.

5.2.5 Second Occurrence Focus

web1 ==> *lab_htk_SOF*

The SOF data (194 tokens) proved considerably more difficult for all of the web-trained classifiers. Impressionistic and experimental studies of second occurrence all agree that, when an SOF occurs following the nuclear pitch accent, as they do in our experimental data, *f0* maxima are reduced or non-existent. As in the case of the interrogative condition, we expect classifiers using quality-related measurements (and in particular *f0* measurements) to perform poorly. Indeed, these classifiers had the weakest performance: 56.2%, 47.4% and 44.8%. Although all of the classifiers using quantity-related features tended to perform relatively poorly on both FOF and SOF data, the SOF data stand out in yielding classifier performance only marginally better, or worse, than the baseline level. Conversely, the best-performing classifiers, at 71.6%, 70.1% and 70.6% generalization accuracies, used feature sets lacking *f0* features: Hand-picked A ('duration_V2', 'duration_C3', 'f1f2Time50_V2'), Best Duration ('duration_V2',

‘duration_syll1’) and Hand-picked D (‘duration_V2’, ‘duration_C3’, ‘f1f2Time40_V2’), respectively.

These results are therefore consistent with previous studies of SOF prosody, which have found that measures of f_0 are not good predictors of SOF. Other acoustic measures, in particular duration, have been shown to correlate with focus weakly, albeit with statistical significance (e.g Beaver *et al.* 2007). That the duration-based classifiers (*viz.* those using the feature set Best Duration) performed best also agrees with previous studies.

		lab_htk_SOF		
Feature set	Baseline	SVM (RBF)	SVM (linear)	LDA
1. Full	51.8	56.7 (30.2)	68.0 (29.3)	--
2. Best Duration (VarSelRF)	51.8	70.6 (21.8)	68.0 (24.0)	66.5 (23.0)
3. Best Quantity (VarSelRF)	51.8	56.2 (41.9)	47.4 (52.1)	44.8 (55.1)
4. Best Quality (VarSelRF)	51.8	63.4 (27.7)	64.4 (27.0)	58.8 (26.3)
5. 2-4 Combined	51.8	61.3 (38.1)	66.5 (30.2)	60.3 (27.6)
6. Best of full (VarSelRF)	51.8	65.5 (34.1)	58.2 (30.3)	58.8 (27.1)
7. Hand-picked A	51.8	69.1 (19.6)	70.1 (19.2)	69.1 (19.6)
8. Hand-picked B	51.8	61.9 (27.7)	58.2 (32.5)	58.8 (27.1)
9. Hand-picked C	51.8	65.5 (28.5)	58.8 (32.1)	59.3 (26.8)
10. Hand-picked D	51.8	69.6 (19.4)	71.6 (19.5)	67.5 (21.4)
11. Best of full (F-score+SVM)	51.8	64.9 (22.5)	69.1 (20.7)	68.0 (21.2)

Table 9. Accuracy rate (and balanced error rate) for different models: training set web1; test set lab_htk_SOF

5.2.6 Discussion

The web-trained classifiers performed competitively on both web data and laboratory data. However, the web-trained classifiers performed better on the web data than on the laboratory data. It is not immediately clear why this should be so.

In principle, the laboratory data should be considerably ‘cleaner’ than the web data, in the sense that they come from controlled, high-quality recordings, which should allow more accurate acoustic analysis. Surprisingly, the error rates for f_0 extraction were similar between the

laboratory-elicited and web-harvested speech. The hand-annotated laboratory data had only 1.8% undefined values for *mean_f0_ratio*, compared with an impressive 1.6% for the web dataset *web2*.

Moreover, using hand annotated laboratory data did not improve classifier performance over the automatically annotated laboratory data, which suggests that the differing performances cannot be attributed to annotation method.

Many of the tokens in the web corpus come from professional radio and television broadcasters, individuals who are known to have clearer and more consistent prosody (e.g. Ostendorf et al. 1995). We can also speculate that the majority of tokens in the web corpus come from less scripted and more spontaneous speech, in contrast to the laboratory speech, which is entirely read speech, elicited in an artificial context. These two factors—the relative naturalness and clear articulation in tokens of web speech and the relative unnaturalness and relaxed articulation in tokens of laboratory speech—likely contribute to better classifier performance on web data.

How much significance we should place on the different classification accuracies is unclear. What is clear, however, is that the results for both test sets—laboratory-elicited and web-harvested—overwhelmingly support the theoretical prediction that prominence in comparative clauses is correlated with the coreference criterion.

The results of all the web-trained classifiers also agree that the available measures of *f0* are at best weakly predictive of focus category. On the web data, the best classifiers with and without *f0* measures performed competitively. On the laboratory data, classifiers without *f0* measures were at a clear advantage. Even more convincing, the performance of classifiers using *f0* measures showed little difference whether the tokens came from declarative or interrogative contexts, suggesting that the non-predictiveness of *f0* measures cannot be attributed to a pooled set of pitch accent types (e.g. high H* and low L*).

Finally, we know independently from other laboratory investigations (e.g. Rooth 1996a, Bartels 2004, Beaver et al. 2007, Bishop 2008, Howell 2010) that second occurrence focus tends to lack phonological pitch accents and is realized with reduced f_0 movement. Advocates of a grammatically mediated theory of focus, who maintain that focus sensitive constructions grammatically require there to be a focus—predict that second occurrence foci must therefore be realized with other, non-intonational prominence. Classifiers using non-intonational measures were strongly robust when tested on regular, first occurrence foci. We expect, then, the classifiers using non-intonational measures to be equally robust when applied to second occurrence focus data. Instead, we find that the performance of these classifiers on SOF data was considerably weaker.

Certain classifiers, in particular those using duration features, performed above the baseline and greater than 70% generalization accuracy. This confirms that SOF may be realized with acoustic prominence. Either the web-trained classifiers are not sufficiently robust, or the acoustic cues for second occurrence focus are less consistent than for first occurrence focus. In Section 5.3.3, I'll apply laboratory-trained classifiers to the second occurrence focus data.

5.3 *Training with laboratory data*

lab_htk_FOF ==>

In this section, we continue the cross-validation of the web and laboratory data by training classifiers on laboratory data and evaluating their performance on web data. I also use web-trained classifiers on the second occurrence focus data elicited in the laboratory.

5.3.1 *Feature selection*

I followed the same procedure for feature selection used in the previous sections: I used the VarSelRF algorithm to automatically generate feature sets and I used the experimenter-selected feature sets from the previous section.

The automatically-generated feature sets differed in constituency, since the algorithm was provided a different dataset. They also differed in size. Many more features were selected, suggesting that individual features may contribute less explanatory power. In several cases, the large number of features meant that some of the features were collinear. An LDA classifier could therefore not be trained using these feature sets (cf. 1,3-6 in Table 11). The feature sets for the laboratory-trained data are summarized in Table 10.

Selection method	Number of features	Features selected
2. Best Duration (VarSelRF)	4	duration_C3, duration_syll1, duration_V2, duration_V3
3. Best Quantity (VarSelRF)	12	energy_ratio, maxf0_ratio, maxf0Time_V3_percent, maxIntensity_ratio, maxIntTime_V3_percent, meanIntensity_ratio, meanIntensity_V3, minf0_V3, minf0Time_V3_percent, minIntTime_V3_percent, power_ratio, power_V3
4. Best Quality (VarSelRF)	27	f1bandTime10_V3, f1f2Time10_V2, f1f2Time20_V2, f1f2Time30_V2, f1f2Time40_V2, f1f2Time50_V2, f1f2Time60_V2, f1f2Time70_V2, f1f2TimeIntmax_V2, f1Time20_V2, f1Time30_V2, f1Time40_V2, f1Time40_V3, f1Time50_V2, f1Time50_V3, f1Time60_V2, f1Time70_V2, f1Time70_V3, f1TimeIntmax_V2, f2Time20_V2, f2Time30_V2, f2Time40_V2, f2Time60_V2, jitter_V3_percent, maxf1_V2, minf2_V2, rangef1_V2
5. 2-4 Combined	43	all of the above
6. Best Overall (VarSelRF)	18	duration_V2, duration_V3, f1f2Time20_V2, f1f2Time30_V2, f1f2Time40_V2, f1f2Time50_V2, f1f2Time60_V2, f1Time30_V2, f1Time40_V2, f1Time50_V2, f1Time60_V2, f2Time30_V2, f2Time40_V2, f2Time50_V2, maxIntensity_ratio, meanIntensity_ratio, minf0Time_V3_percent, power_ratio
7. Hand-picked A	3	duration_V2, duration_C3, f1f2Time50_V2
8. Hand-picked B	5	duration_V2, duration_C3, mean_f0_ratio, f1f2Time40_V2, f1f2Time50_V2
9. Hand-picked C	4	duration_V2, duration_C3, mean_f0_ratio, f1f2Time50_V2
10. Hand-picked D	3	duration_V2, duration_C3, f1f2Time40_V2
11. Best overall (F-score+SVM)	8	h1minusa2F1_V1, f2Time30_V3, rangef2_V2, maxIntensity_ratio, maxf0_V3, jitter_V2_percent, duration_syll1, duration_V2

Table 10. List of feature sets selected for training on dataset lab_htk_FOF

5.3.2 Classification accuracy on web-harvested data

lab_htk_FOF ==> *web1+2*

Nearly all of the laboratory-trained classifiers performed competitively and all of them performed above the baseline. The worst-performing classifiers used quantity-related features; above we saw that classifiers using quantity-related features also performed poorly when trained on web data. The best performing classifier trained on the laboratory data used experimenter-selected feature set Hand-picked B ('duration_V2', 'duration_C3', 'mean_f0_ratio', 'f1f2Time40_V2', 'f1f2Time50_V2'). This classifier achieved a rate of accuracy on par with the best web-trained classifier: 90.4% and 92.9%, respectively.

Feature set	<i>web_1+2: web1 & web2 (web corpus)</i>			
	Baseline	SVM (RBF)	SVM (linear)	LDA
1. Full	50.5	79.8 (17.4)	73.4 (24.7)	--
2. Best Duration (VarSelRF)	50.5	82.1 (17.9)	82.6 (17.2)	81.7 (18.2)
3. Best Quantity (VarSelRF)	50.5	62.4 (36.7)	60.1 (39.9)	--
4. Best Quality (VarSelRF)	50.5	81.2 (17.9)	75.2 (20.9)	--
5. 2-4 Combined	50.5	83.9 (15.2)	79.4 (20.6)	--
6. Best of full (VarSelRF)	50.5	81.7 (16.8)	72.9 (27.1)	--
7. Hand-picked A	50.5	85.8 (12.9)	88.5 (10.3)	88.1 (10.9)
8. Hand-picked B	50.5	90.4 (9.2)	89.0 (10.0)	89.4 (10.0)
9. Hand-picked C	50.5	89.4 (9.8)	88.5 (10.3)	88.1 (10.9)
10. Hand-picked D	50.5	84.9 (14.1)	89.9 (9.3)	89.9 (9.6)
11. Best of full (F-score+SVM)	50.5	80.7 (16.8)	78.0 (18.2)	81.7 (17)

Table 11. Accuracy rate (and balanced error rate) for different models: training set web1; test set lab_htk_FOF

5.3.3 Second occurrence focus

lab_htk_FOF ==> *lab_htk_SOF*

The laboratory-trained classifiers were trained on the regular, first occurrence focus data. The second occurrence focus data, elicited from the same set of speakers, constituted a distinct, though related, dataset. Recall that the web-trained classifiers underperformed on the SOF data, in comparison to the FOF data (Section 5.2.5). The top-performing web-trained classifier

achieved an accuracy rate of 71.6% (19.5% BER). In contrast, the top-performing lab-trained classifier achieved an accuracy rate of 81.0% (18.1% BER). This classifier was a linear SVM using the feature set Hand-picked D (‘duration_V2’, ‘duration_C3’, ‘flf2Time40_V2’).

		<i>lab_htk_SOF</i>		
Feature set	Baseline	SVM (RBF)	SVM (linear)	LDA
1. Full	51.0	76.4 (19.3)	74.9 (27.4)	--
2. Best Duration (VarSelRF)	51.0	75.4 (24.4)	76.9 (22.3)	74.9 (23.4)
3. Best Quantity (VarSelRF)	51.0	66.7 (26.5)	67.2 (29.6)	--
4. Best Quality (VarSelRF)	51.0	80.0 (19.4)	78.5 (21.2)	--
5. 2-4 Combined	51.0	80.5 (18.3)	76.4 (23.2)	--
6. Best of full (VarSelRF)	51.0	80.5 (19.4)	76.9 (23.0)	--
7. Hand-picked A	51.0	80.5 (18.0)	80.0 (18.7)	77.4 (20.2)
8. Hand-picked B	51.0	79.5 (20.5)	79.0 (20.0)	78.5 (19.5)
9. Hand-picked C	51.0	79.5 (20.5)	80.0 (18.7)	77.4 (20.2)
10. Hand-picked D	51.0	80.5 (18.5)	81.0 (18.1)	79.0 (18.8)
11. Best of full (F-score+SVM)	51.0	75.4 (24.2)	77.4 (22.6)	75.4 (24.3)

Table 12. Accuracy rate (and balanced error rate) for different models: training set web1; test set lab_htk_FOF

5.3.4 Discussion

The high generalization accuracies of the laboratory-trained classifiers suggest that the efficacy of the web and laboratory data for training is somewhat asymmetrical. Both laboratory and web speech may be used as training data for web speech. However, the web speech proved less effective as training data for laboratory speech.

This asymmetry is consistent with the observation that many of the tokens from the web dataset were produced by professional broadcasters. They are less likely to produce speech that is potentially ambiguous (e.g. produced with coarticulation and reduced or epenthized vowels) and more likely to mark prosody consistently (Ostendorf et al. 1995) and with hyperarticulation. It follows then, that a classifier trained on hyperarticulated speech would have more difficulty when applied to laboratory speech than on similarly hyperarticulated speech; and a classifier

trained on laboratory speech would do equally well on both hyperarticulated and laboratory speech.

As a consequence, the web data may offer an important source of cross-validation, not only because they are produced more naturally and in a variety of different pragmatic contexts, but because the web speech appears to contain tokens with more idealized realizations.

The laboratory-trained classifiers, trained on FOF data, also performed well on the laboratory-elicited SOF dataset. They performed competitively with other classifiers in this study and much better than the web-trained classifier. These results for SOF are more robust than one might expect from previous investigations, which demonstrated a measurable, statistically significant acoustic effect, but of questionable magnitude. Using data from the same set of speakers for training and testing likely improved classifier performance, particularly if it is the case that acoustic cues of prominence vary with speaker (e.g. Mo 2011). We also cannot be certain that all of the second occurrence focus data, elicited in the laboratory, were produced naturalistically, with appropriate prosody. Nonetheless, the success of the classifiers could not be possible without robust acoustic cues of prominence. If we allow that the SOF tokens were produced naturalistically, the results would confirm that speakers can and do realize second occurrence focus with non-intonational prominence. Moreover, the classifiers using non-intonational acoustic measures performed well on both FOF and SOF; this would suggest that SOF is not realized with different cues from FOF.

6 *Human acoustic classifiers*

In this section, I assess the validity of the machine learning classifier results by comparing the machine learning classifiers to human classifiers. In other words, we want to know how closely the machine learning classifiers mimic human speech perception in classification accuracy and the acoustic measurements used to make judgements. I conducted two perceptual

experiments to answer this question: the first using stimuli from the web dataset; the second using stimuli from the laboratory dataset.

6.1 Experiment 1: web stimuli

6.1.1 Methodology

A subset of 64 tokens from the *web2* corpus dataset was chosen: the first 32 of each semantic focus class. From each soundfile, the sequence “than I did” was extracted to create the stimuli. The files were normalized for sampling frequency and amplitude. The information presented to participants of the perception experiment was limited in this way in order to more closely match the limited information available to the statistical classifiers: neither machine nor human had the preceding or following acoustic information and neither machine nor human had any linguistic or extra-linguistic context.

Forty individuals participated in the perception experiment, which was conducted at the prosody lab at McGill University. Participants were compensated for their time. The data of two participants was not analyzed because the subjects reported making errors. The stimuli were played one at a time, in random order, with no category repeated more than twice. After each stimuli, the listener was asked to complete two tasks: first, to choose whether “I” or “did” had greater prominence; second, to rate confidence in their choice on a scale from 1 (“very confident”) to 5 (“very uncertain”).

(14) Questions elicited in laboratory perception experiment

Question 1:	Which is more prominent: I or did?
Question 2:	How confident are you?
	(very uncertain) 1 2 3 4 5 6 7 (very confident)

I evaluate the results in two ways. First, I calculate accuracy rates and balanced error rates, just as I did for the machine learning classifiers. In this way we can compare the human and machine learning classifiers using the same performance measures. We can also compare these

measures by listener and by item. If many listeners or many items are misclassified consistently, this would suggest a listener or item bias.

Second, I evaluate generalized linear mixed models using two of the top-performing feature sets. Mixed models allow us to incorporate, and essentially filter out, random effects of listener and item. I chose the two feature sets *Hand-picked A* ('duration_V2', 'duration_C3', 'f1f2Time50_V2') and *Hand-picked C* ('duration_V2', 'duration_C3', 'f1f2Time50_V2', 'mean_f0_ratio'), because they were used for most of the top-performing machine learning classifiers, and because they differed in a single feature of interest, namely *mean_f0_ratio*. The modeling allows us to ask two related but independent questions: (1) how much variance in *listeners' responses* does a model using these features predict; and (2) how much variance in *correct responses* does a model using these features predict.

For a given model, we can also ask how predictive the individual features in the model are, and whether the model predicts significantly more variation than another model.

6.1.2 Results

Accuracy/Error

As a group, the 38 participants achieved a mean accuracy of 85.9%, median accuracy of 89.1% and standard deviation of accuracies 8.3%. They achieved a mean BER of 14.1%, median BER of 10.9% and standard deviation of BERs 8.3%. Participants' individual accuracy rates ranged from 64.1% to 95.3% and their balanced error rates ranged from 4.7% to 35.9% percent. These distributions are illustrated in Figures 20 and 21.

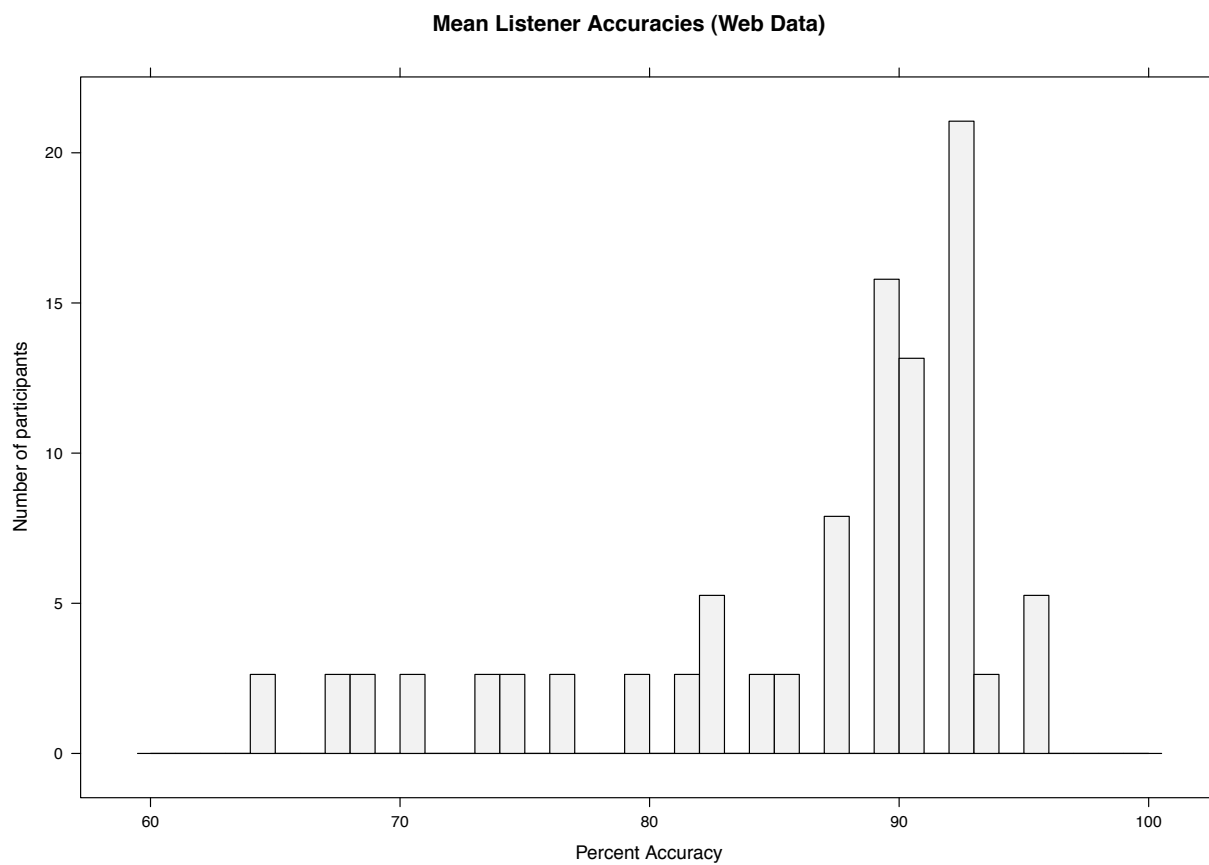


Figure 20. Distributions of listener accuracies (n=38): mean = 85.9%, median=89.1%, stdev=8.3%.

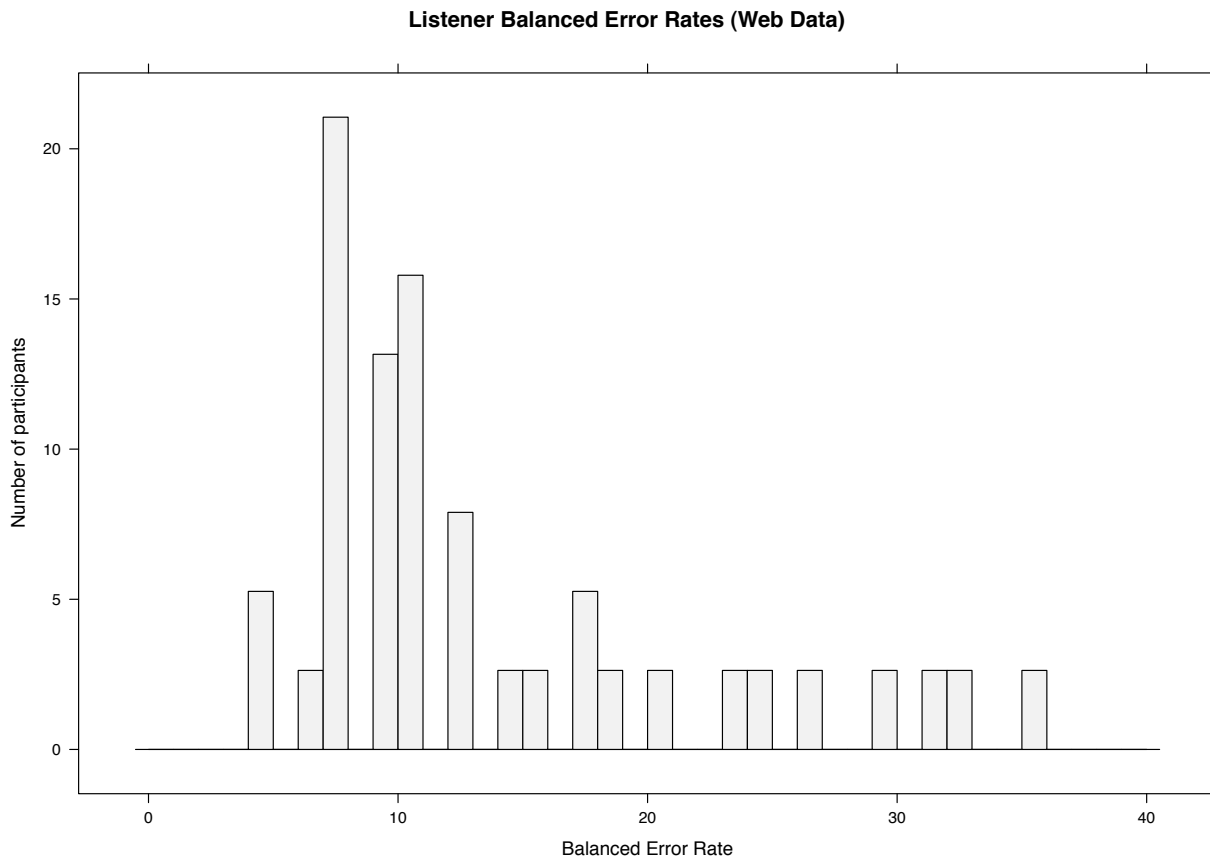


Figure 21. Distributions of listener balanced error rates (n=38): mean = 14.1%, median 10.9%, stdev=8.3%.

As for the items used in the experiment, only 3 were consistently misidentified by listeners. The majority of the stimuli were classified correctly more than 80% of the time (cf. Figure 22). The mean by item accuracy rate was 85.9%, the median 89.5% and the standard deviation 16.9%.

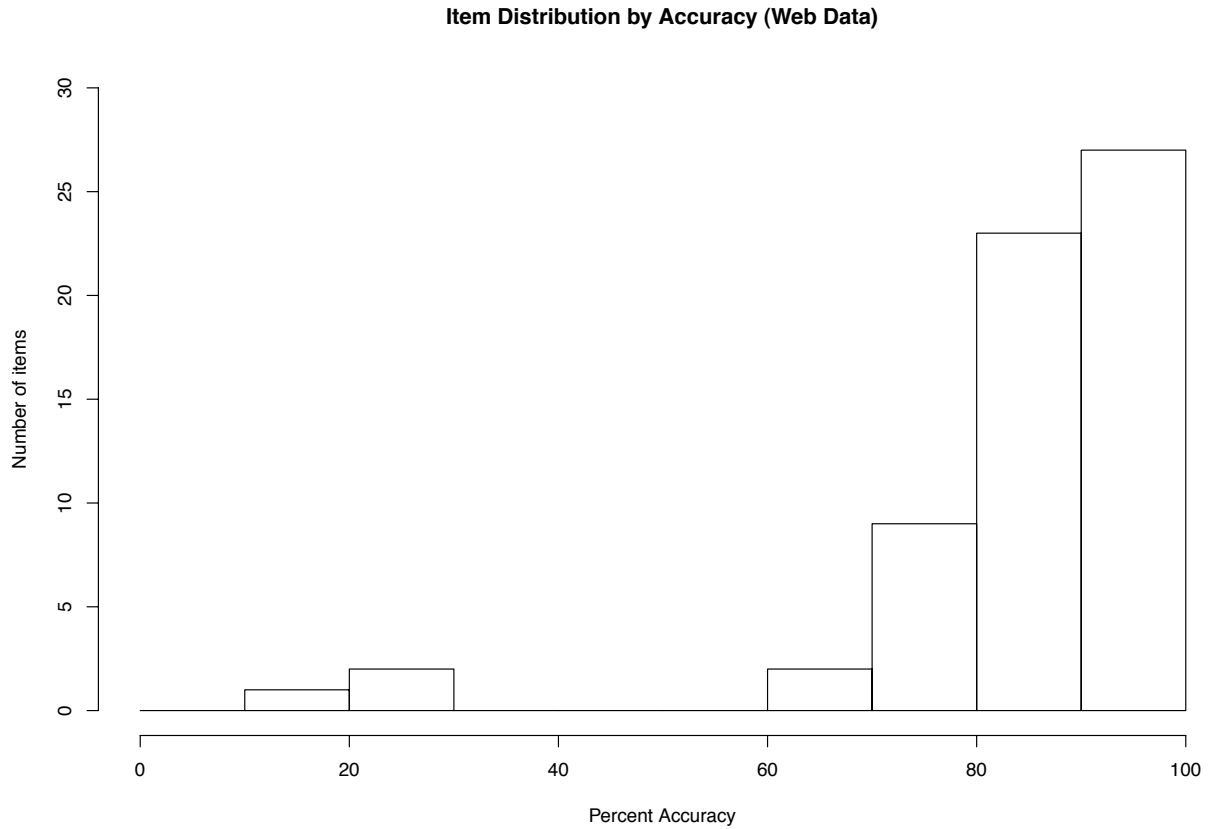


Figure 22. Distribution of items (n=64) from *web2* according to listener accuracy rate. Mean=85.9%; median=89.5%; standard deviation=16.9%.

Generalized Linear Mixed Models

In order to understand which acoustic features listeners were using to make their judgments, I tested for the statistical significance of individual features in “generalized linear mixed models” using the R package lme4 (Bates, Maechler & Bolker 2007). I considered one model which had the featureset *Hand-picked A* (features ‘duration_V2’, ‘duration_C3’ and ‘f1f2Time50_V2’) as fixed effects; and I considered a second model which had the featureset *Hand-picked C* (features ‘duration_V2’, ‘duration_C3’, ‘f1f2Time50_V2’ and ‘mean_f0_ratio’). The two models differed only in the feature ‘mean_f0_ratio’. Both models also contained participant and item as random effects.

Both statistical models were significant, as were each of the individual fixed effects (i.e. the acoustic features), with the notable exception of *mean_f0_ratio* (cf. Table 13).

Generalized Linear Mixed Model of Listener Response (Web Data)

HAND-PICKED A: duration_V2, duration_C3, flf2Time50_V2

Random effects:

Groups	Variance	Std. Dev.
Participant	0.066720	0.25830
Item	0.041699	0.20420

Fixed effects:

	Estimate	Std. Error	z-value	p-value
<i>Intercept</i>	1.210236	0.520745	2.324	0.0201 *
duration_V2	35.946678	2.254716	15.943	<2e-16 *
duration_C3	-45.078265	3.401762	-13.251	<2e-16 *
flf2Time50_V2	-0.003068	0.000329	-9.326	<2e-16 *

HAND-PICKED C: duration_V2, duration_C3, flf2Time50_V2, mean_f0_ratio

Random effects:

Groups	Variance	Std. Dev.
Participant	0.066720	0.25830
Item	0.041699	0.20420

Fixed effects:

	Estimate	Std. Error	z-value	p-value
<i>Intercept</i>	1.249e+00	5.746e-01	2.174	0.0297 *
duration_V2	3.605e+01	2.332e+00	15.457	<2e-16 *
duration_C3	-4.524e+01	3.509e+00	-12.893	<2e-16 *
flf2Time50_V2	-3.067e-03	3.291e-04	-9.318	<2e-16 *
mean_f0_ratio	-4.232e-02	2.654e-01	-0.159	0.873 n.s.

Table 13. Summary of generalized linear mixed models for listener responses to a subset of web2 using predictors from hand-selected feature sets Hand-picked A and Hand-picked C. Test statistic Wald z-score; statistical significance ($p < 0.01$) indicated by asterisks.

We can quantify whether one the two models of listener response is more predictive than the other using an Analysis of Variance (ANOVA) (cf. Baayen, Davidson & Bates 2008). The various goodness of fit criteria (AIC, BIC and log likelihood) for our two models are very similar and according to the χ^2 test statistic, we cannot conclude that the model using feature set *Experimenter-Selected C* predicts significantly more variation than the model using feature set *Experimenter-Selected A*. In other words, we cannot say that adding the feature *mean_f0_ratio* results in a more predictive model of listeners' responses.

	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)	
Random: Participant, Item Fixed: <i>Hand-picked A</i>	6	1741.19	1775.96	-864.59				
Random: Participant, Item Fixed: <i>Hand-picked C</i>	7	1743.28	1783.85	-864.64	0	1	1	n.s.

Table 14. Summary of Analysis of Variance (ANOVA) between the two mixed-effect models in table 13. Df is the degrees of freedom in the model; AIC (Akaike information criterion), BIC (Bayesian information criterion) and log likelihood measure goodness of fit of a model. The test statistic used is χ^2 ; p-value indicates value of statistical significance.

We can also perform model comparison to assess the contribution of the random effects, participant and item. The ANOVA in Table 15 shows that a model with item as random effect explains significantly more variance than a model without it; the ANOVA in Table 16 shows that a model with participant as random effect explains significantly more variance than a model without it.

	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)	
Random: Participant, Item Fixed: <i>Hand-picked A</i>	5	1819.72	1848.70	-864.59				
Random: Participant Fixed: <i>Hand-picked A</i>	6	1741.19	1775.96	-864.59	80.533	1	< 2.2e-16	*

Table 15. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models (fixed effects from featureset *Hand-picked A*) with and without item as random effect.

	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)	
Random: Participant, Item Fixed: <i>Hand-picked A</i>	5	1760.65	1789.63	-875.33				
Random: Item Fixed: <i>Hand-picked A</i>	6	1741.19	1775.96	-864.59	21.465	1	3.603e-16	*

Table 16. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models (fixed effects from featureset Hand-picked A) with and without participant as random effect.

Confidence Rating

Participants' confidence rating turned out to be a very significant predictor of their performance on a given stimuli (generalized linear model: $\sigma = 0.031$, $z = -10.81$, $p < 0.001$). This suggests that, when they performed well, listeners were not simply guessing.

6.1.3 Discussion

The performance of listeners in the perception experiment, as measured by classification accuracy and balanced error rate, closely matched that of the machine learning classifiers. Recall that the top-performing classifier achieved an accuracy rate of 92.9%: while some listeners' accuracy rates were as low as 64%, 16 out of the 38 human classifiers achieved an accuracy rate above 90%.

The comparison of listener response models revealed that item explains a statistically significant amount of listener variation. Review of the item distribution in Figure 22, however, reveals that 3 of the 64 items were effectively outliers, with accuracy rates well below 50%. The poor human classifier performance on these items suggests that misclassification by the machine learning classifiers are likely to be a result of other variation (e.g. speaker disfluency, high signal-to-noise ratio) in the data by which human listeners were equally misled.

One misclassified example, transcribed in (15), received a listener accuracy rate of 18.4%. The co-reference criterion predicts this example will be realized with subject focus since the

subjects of the two clauses do not co-refer; however, the matrix clause also has a salient contrast ‘at that time’ which licenses focus on *did* as well.

- (15) Growing up at that time and that location, you can’t have more fun as a kid
than [I]_F [did]_F

In other words, (15) is an infrequently occurring but linguistically possible example of double focus. The task of the machine learning classifiers and the human listeners was binary (two semantic classes for the machine learning classifiers and two prominence choices for the human listeners), whereas (15) belongs to a third class.

Finally, the same featuresets used in the top-performing machine learning classifiers (viz. *Hand-picked A* and *Hand-picked C*) were statistically significant in a model of listener response. There was no main effect for the mean f0 ratio feature (i.e. it was not individually significant in the model), and removing the feature did not result in a less explanatory model. This result is consistent with the corresponding machine learning classifiers, for which the addition of the feature *mean_f0_ratio* did not noticeably improve generalization accuracy or error rates.

6.2 Experiment 2: laboratory production stimuli

6.2.1 Methodology

In the second perception experiment, human listeners were presented with excerpts of “than I did” taken from a subset of the laboratory production data.¹⁷ The experiment was carried out with the same methodology as in Experiment 1. Forty-one individuals participated.

¹⁷ I used speech from the first 8 participants of the production study. I used 8 of the original 16 elicited utterances—the same 8 for each of the 8 speakers: tokens 1, 3, 5, 7, 9, 11, 13 and 15.

6.2.2 Results

Accuracy/Error

The human acoustic classifiers performed on par with the machine learning classifiers. For the FOF data, the 41 listeners achieved a mean accuracy of 78.5% , median accuracy 81.3% and standard deviation of accuracies 13.1%. They achieved a mean balanced error rate of 13.1%, median balanced error rate of 12.2% and a standard deviation of balanced error rates 6.9%. Participants' individual accuracy rates ranged from 53.1% to 96.9% and their balanced error rates ranged from 3.7% to 29.3%. The distributions of individual accuracy rates and balanced error rates for FOF are illustrated in Figures 23 and 24.

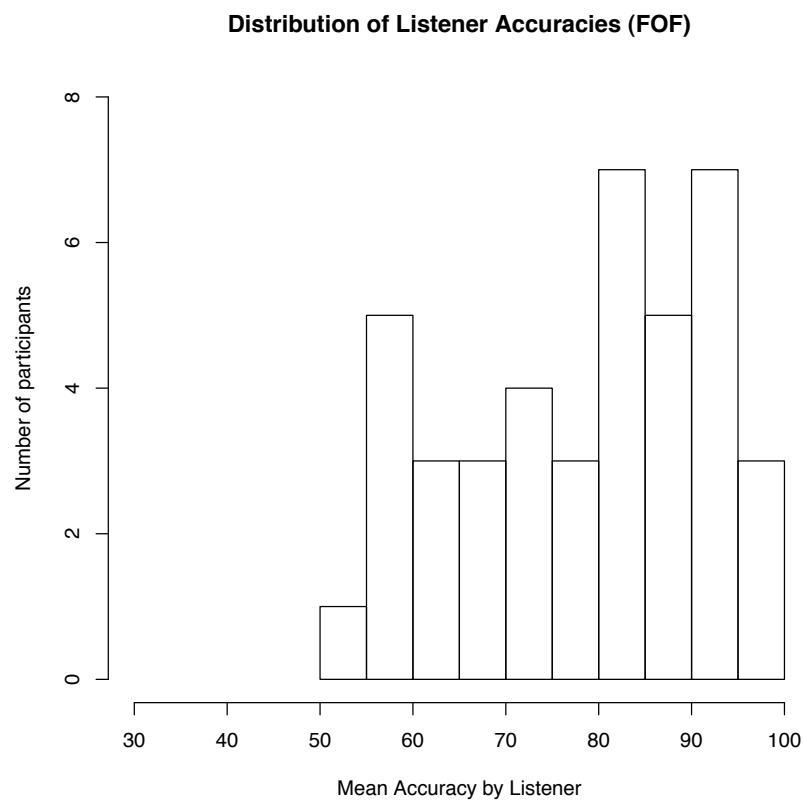


Figure 23. Distribution of listener accuracies. First occurrence focus: Mean= 78.5%, Median=81.3%, Stdev=13.1

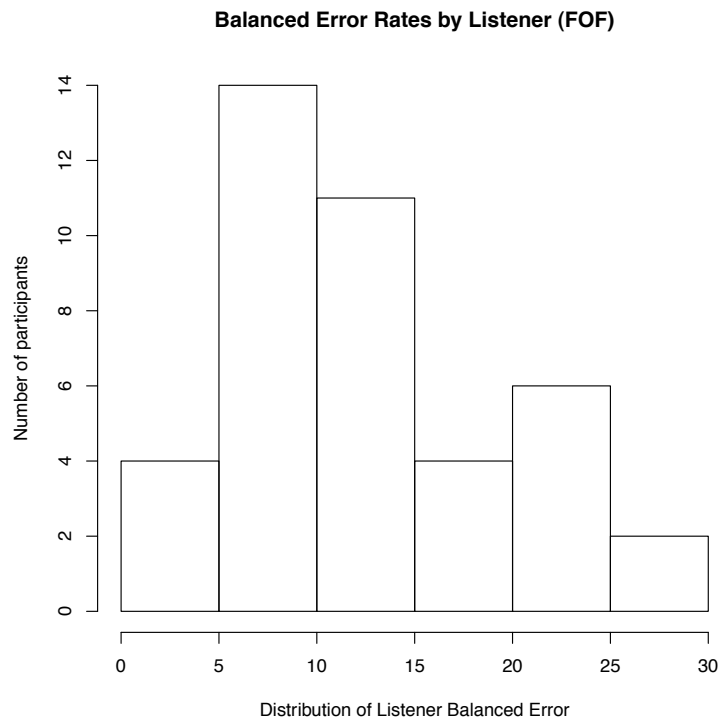


Figure 24. Distribution of participants (n=41) according to the participant balanced error rate. First occurrence focus: Mean=13.1%, Median=12.2%, Stdev=6.9%.

For the second occurrence focus data, the 41 listeners achieved a mean accuracy of 71.2%, median accuracy 75.0% and standard deviation of accuracies 15.6%. They achieved a mean balanced error rate of 15.8%, median balanced error rate of 15.9% and a standard deviation of balanced error rates 6.6%. Participants' individual accuracy rates ranged from 34.4% to 93.8% and their balanced error rates ranged from 6.1% to 34.2%. The distributions of individual accuracy rates and balanced error rates for are illustrated in Figures 25 and 26.

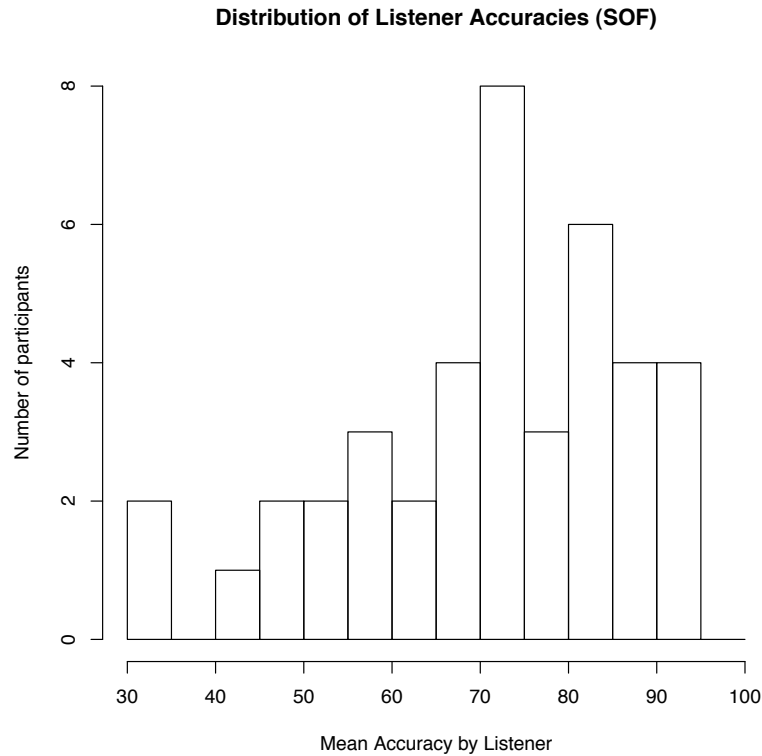


Figure 25 Distribution of listeners (n=41) according to the listener accuracy rate. Second occurrence focus: Mean= 71.2, Median=75.0, Stdev=15.6

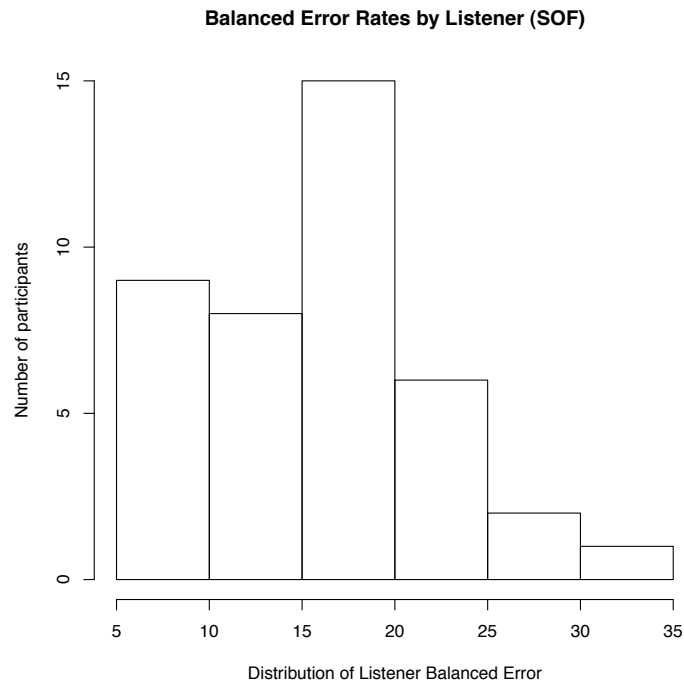


Figure 26. Distribution of participants (n=41) according to the participant balanced error rate. Second occurrence focus: Mean=15.8, Median=15.9, Stdev=6.6.

As for the items used in the experiment, 1 FOF item was correctly identified at less than 50% and 3 SOF items were correctly identified at less than 50%. Among FOF items, the mean accuracy rate was 78.5%, median 80.5% and the standard deviation 16.9%. Among SOF items the mean accuracy rate was 71.2%, median 72.0% and standard deviation 15.1%.

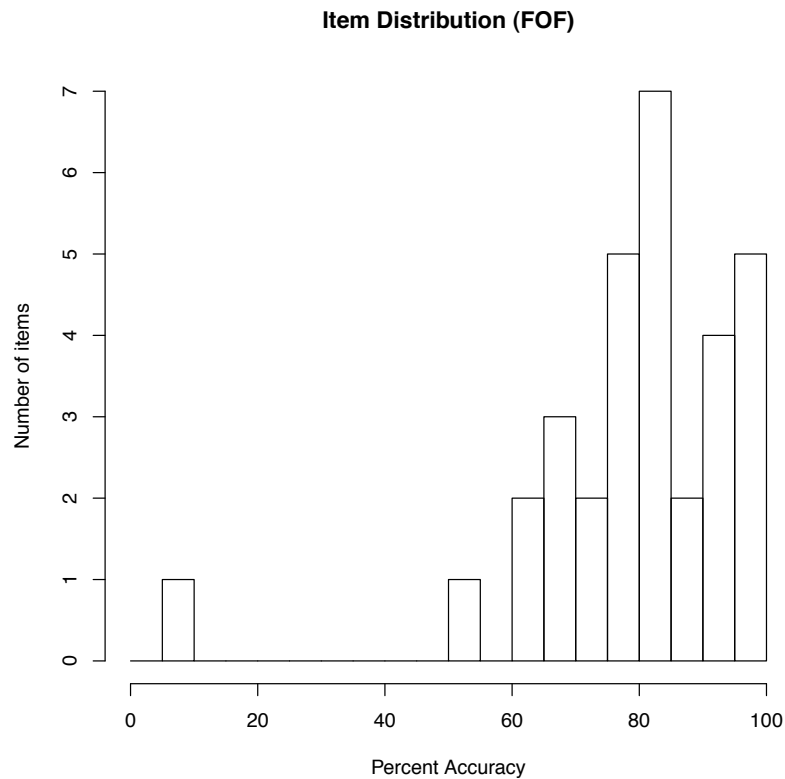


Figure 27. Distribution of items from laboratory production experiment according to listener accuracy rate. First occurrence focus(n=32). Mean=78.5%; Median=80.5%; Stdev=16.9%.

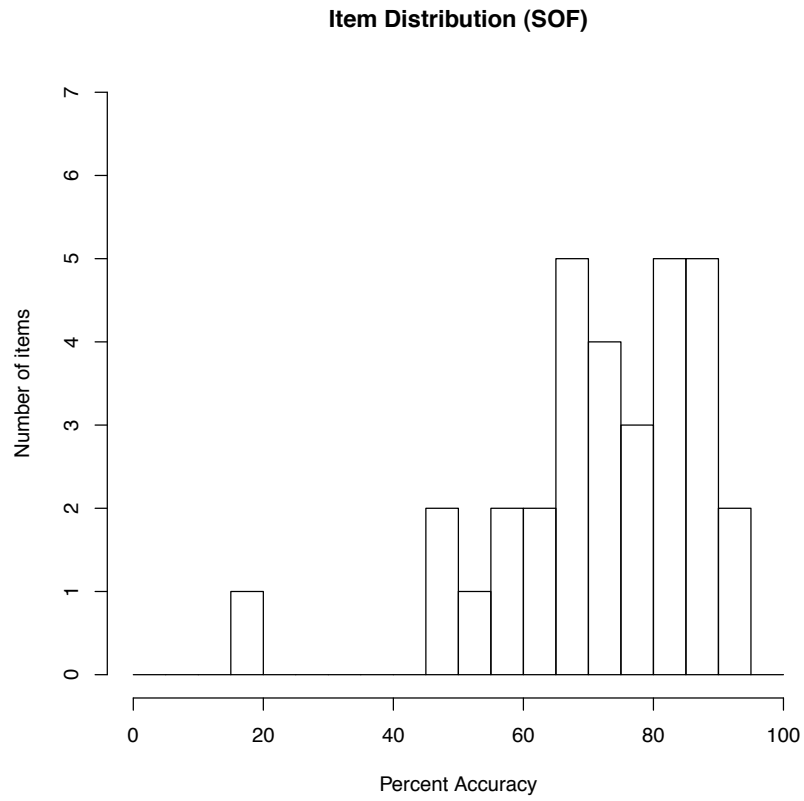


Figure 28. Distribution of items from laboratory production experiment according to listener accuracy rate.. Second occurrence focus (n=32). Mean=71.2%; Median=72.0%; Stdev=15.1%.

Stimuli were drawn from eight different speakers in the production experiment. The accuracy rates for individual speakers ranged from 67.4 to 82.0%. The mean accuracy rate among speakers was 74.8%, median 73.9% and standard deviation 5.0%.

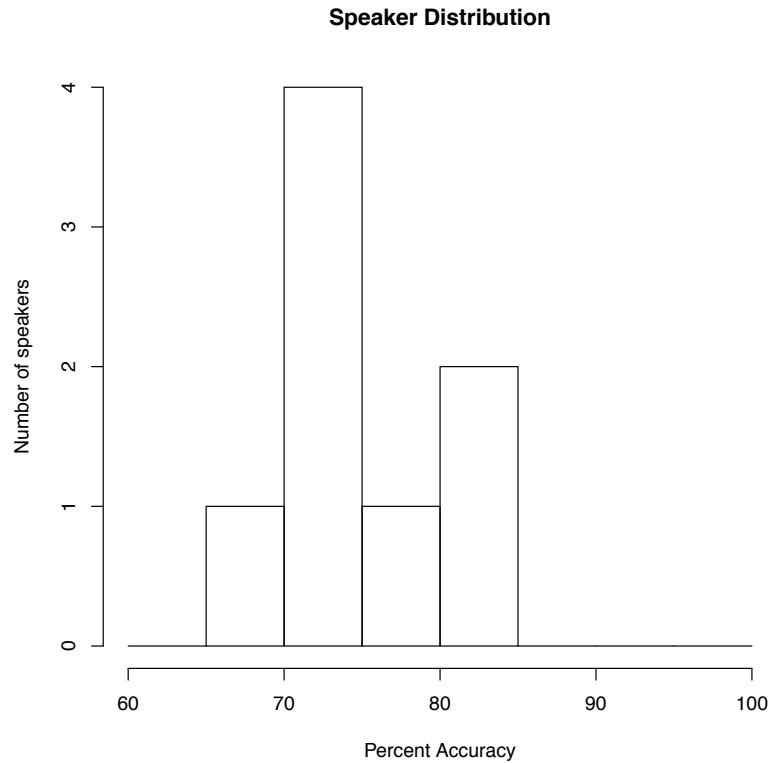


Figure 29. Distribution of speakers (n=8) from laboratory production experiment according to listener accuracy rate: mean=74.8%, median=73.9%, Stdev=5.0%.

Generalized Linear Mixed Models

As in the first perception experiment, in order to understand which acoustic features listeners were using to make their judgments, I evaluated generalized linear mixed models using two top-performing feature sets. *Hand-picked A* contains the features ‘duration_V2’, ‘duration_C3’ and ‘f1f2Time50_V2’ as fixed effects; *Hand-picked* contains the features ‘duration_V2’, ‘duration_C3’, ‘f1f2Time50_V2’ and ‘mean_f0_ratio’ as fixed effects. The two models differed only in the feature ‘mean_f0_ratio’. Both models also contained participant and item as random effects. I evaluated separate models for the FOF and SOF data.

All of the listener response models, both for FOF and SOF data, were statistically significant. There were main effects for each of the acoustic features, with the notable exception of ‘mean_f0_ratio’, which was not significant in either the FOF or SOF models. The

feature ‘duration_V3’ was only marginally significant in the FOF model using feature set *Hand-picked A*.

Generalized Linear Mixed Model of Listener Response (Lab Data: FOF)

HAND-PICKED A: duration_V2, duration_C3, flf2Time50_V2

Random effects:

Groups	Variance	Std. Dev.
Participant	0.08346	0.28890
Item	0.06514	0.25523

Fixed effects:

	Estimate	Std. Error	z-value	p-value
<i>Intercept</i>	5.261e+00	1.209e+00	4.35	1.36e-05 *
duration_V2	1.398e+01	4.203e+00	3.327	0.00088 *
duration_C3	-1.742e+01	9.574e+00	-1.820	0.06879 n.s.
flf2Time50_V2	-5.719e-03	6.446e-04	-8.872	< 2e-16 *

HAND-PICKED C: duration_V2, duration_C3, flf2Time50_V2, mean_f0_ratio

Random effects:

Groups	Variance	Std. Dev.
Participant	0.42816	0.65434
Item	0.61338	0.78318

Fixed effects:

	Estimate	Std. Error	z-value	p-value
<i>Intercept</i>	6.022606	2.128936	2.829	0.00467 *
duration_V2	16.005712	7.147043	2.239	0.02512 *
duration_C3	-20.864184	18.070114	-1.155	0.24824 n.s.
flf2Time50_V2	-0.006340	0.001154	-5.493	3.95e-08 *
f0_ratio	0.352155	0.298404	1.180	0.23795 n.s.

Table 17. Summary of generalized linear mixed models for listener responses to a subset of lab_SOF using predictors from hand-selected feature sets Hand-picked A and Hand-picked C. Test statistic Wald z-score; statistical significance ($p < 0.01$) indicated by asterisks.

Generalized Linear Mixed Model of Listener Response (Lab Data: SOF)

HAND-PICKED A: duration_V2, duration_C3, flf2Time50_V2

Random effects:

Groups	Variance	Std. Dev.
Participant	0.083524	0.28901
Item	0.065190	0.25532

Fixed effects:

	Estimate	Std. Error	z-value	p-value	
<i>Intercept</i>	3.050e+00	1.196e+00	2.55	0.0108	*
duration_V2	1.617e+01	3.995e+00	4.048	5.16e-05	*
duration_C3	-1.388e+01	6.144e+00	-2.259	0.0239	*
flf2Time50_V2	-3.946e-03	6.816e-04	-5.789	7.07e-09	*

HAND-PICKED C: duration_V2, duration_C3, flf2Time50_V2, mean_f0_ratio

Random effects:

Groups	Variance	Std. Dev.
Participant	0.42816	0.65434
Item	0.61338	0.78318

Fixed effects:

	Estimate	Std. Error	z-value	p-value	
<i>Intercept</i>	3.053e+00	1.197e+00	2.551	0.0107	*
duration_V2	1.62e+01	4.038e+00	4.011	6.04e-05	*
duration_C3	-1.391e+01	6.18e+00	-2.25	0.0244	*
flf2Time50_V2	-3.946e-03	6.818e-04	-5.787	7.15e-09	*
f0_ratio	-5.69e-03	1.342e-01	-0.042	0.9662	n.s.

Table 18. Summary of generalized linear mixed models for listener responses to a subset of lab_SOF using predictors from hand-selected feature sets Hand-picked A and Hand-picked C. Test statistic Wald z-score; statistical significance ($p < 0.01$) indicated by asterisks.

Despite a marginal test statistic for the ‘f0_ratio’ parameter estimate in the FOF model ($p=0.23795$ is small but above an acceptable rate of $\alpha=0.05$), an ANOVA comparing the two FOF models without and without ‘f0_ratio’ suggests that the addition of this feature does indeed result in a more predictive model of listener response.

FOF Data	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)	
Random: Participant, Item Fixed: <i>Hand-picked A</i>	6	1819.72	1848.70	-864.59				
Random: Participant, Item Fixed: <i>Hand-picked C</i>	7	1741.19	1775.96	-864.59	80.533	1	< 2.2e-16	*

Table 19. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models for FOF data (cf. Table 17).

Unsurprisingly, an ANOVA comparing the two models of SOF data with and without ‘f0_ratio’ does not suggest that the addition of this feature results in a more predictive model. The various goodness of fit criteria (AIC, BIC and log likelihood) for the two models are very similar; according to the χ^2 test statistic, the two models are not significantly different.

SOF Data	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)	
Random: Participant, Item Fixed: <i>Hand-picked A</i>	6	1515.88	1546.94	-751.94				
Random: Participant, Item Fixed: <i>Hand-picked C</i>	7	1517.89	1554.13	-751.95	0	1	1	n.s.

Table 20. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models for SOF data (cf. Table 18).

Finally, we want to compare the models with and without participant and item as random effects. ANOVAs revealed that the addition of item as random effect resulted in a more predictive model of listener response on FOF and SOF data (cf. Tables 21 & 23). There was not sufficient evidence to conclude that including participant as random effect resulted in more predictive model.

FOF Data	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)	
Random: Participant, Item Fixed: <i>Experimenter-Selected A</i>	5	1232.04	1257.93	-611.02				
Random: Item Fixed: <i>Experimenter-Selected A</i>	6	1261.76	1292.83	-624.88	0	1	1	n.s.

Table 21. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models (fixed effects from feature set Hand-picked A) for FOF data with and without participant as random effect.

FOF Data	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)	
Random: Participant, Item Fixed: <i>Hand-picked A</i>	5	1273.4	1299.31	-631.7				
Random: Participant Fixed: <i>Hand-picked A</i>	6	1261.8	1554.1	-624.9	13.654	1	0.0002198	*

Table 22. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models (fixed effects from feature set Hand-picked A) for FOF data with and without item as random effect.

SOF Data	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)	
Random: Participant, Item Fixed: <i>Hand-picked A</i>	5	1515.35	1541.24	-752.68				
Random: Item Fixed: <i>Hand-picked A</i>	6	1515.88	1546.94	-751.94	1.468	1	0.2256	n.s.

Table 23. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models (fixed effects from feature set Hand-picked A) for SOF data with and without participant as random effect.

SOF Data	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)	
Random: Participant, Item Fixed: <i>Hand-picked A</i>	5	1537.04	1562.93	-763.52				
Random: Participant Fixed: <i>Hand-picked A</i>	6	1515.88	1546.94	-751.94	23.158	1	1.492e-06	*

Table 24. Summary of Analysis of Variance (ANOVA) comparing the two mixed-effect models (fixed effects from feature set Hand-picked A) for SOF data with and without item as random effect.

Confidence Rating

Participants' confidence rating turned out to be a very significant predictor of their performance on a given stimulus (generalized mixed-effects linear model: FOF $\sigma = 0.05844$, $z = 7.429$, $p < 1.10e-13$; SOF $\sigma = 0.04992$, $z = 7.089$, $p = 1.35e-12$; pooled $\sigma = 0.03846$, $z = 10.274$, $p < 2e-16$). This suggests that, when they performed well, listeners were not simply guessing.

6.2.3 Discussion

The performance of listeners in the perception experiment, as measured by classification accuracy and balanced error rate, was on par with that of the machine learning classifiers, although the SOF data proved challenging for both human and machine classifiers. Listeners performed better on FOF data than on SOF data, and listener performance was much more variable on the SOF data: listener performance ranged from 34.4% to 93.8% accuracy and 6.1% to 34.2% BER.

On average, the machine learning classifiers outperformed the human listeners on SOF data: 81.0% accuracy and 18.1% BER for the best-performing classifier and an average 71.2% accuracy and 15.8% BER for listeners. Many listeners, 14 out of the 41 listeners, outperformed the top classifier. Since confidence rating was a significant predictor of performance, and model comparison did not reveal participant to be a predictive feature in a listener response model, it seems unlikely these high-performing listeners were merely fortunate guessers.

In general, the results of this perceptual experiment appear to confirm that speakers can and do realize second occurrence focus with prominence: there must be information present in the signal in order for listeners to achieve as high as 93.8% accuracy. Moreover, the statistical significance of the listener response models suggest that listeners are using the same acoustic features as the machine learning classifiers.

This interpretation of the SOF results, however, rests on the not uncontroversial assumption that participants in the production study produced genuine naturalistic utterances of

SOF. What remains to be explained is the increased variation in listener accuracy and the robust but consistently poorer performances of human and machine learning classifiers on SOF data. The variation and degraded performance are consistent with a combination of both genuine SOF and erroneous SOF (i.e. FOF), with the former having the effect of increasing variance and lower mean and median accuracy values.

Unfortunately, SOF is too rare and elusive a phenomenon to find significant numbers of, even in a vast corpus of web speech. In Howell (2010), I attempt to elicit naturalistic speech by using non-naïve speakers; the results suggested that SOF was not realized consistently with robust phonetic cues. Another experimental approach would be to elicit SOF as part of real-world tasks and to test for speaker comprehension. I leave this for future research.

In addition to the possibility of unnatural conditions for production of SOF, we must also consider the possibility of unnatural conditions of perception. The humans were given a similar task to the machine learning classifiers, i.e. they were provided only with acoustic information for part of an utterance, but yet task may not reflect naturally occurring perception of SOF. First, listeners compared the prominence of two words *within* an utterance; as discussed in section 7.3, it may be the case that listeners compare both *within* and utterance and *across* possible utterances in a natural speech situation. Second, the task is necessarily artificial: we are never required to compare the prominence of two words out of context in a natural speech situation; some listeners may be better at the more analytical, metalinguistic task.

As for the machine learning classifiers, they were provided with an imperfect and often incomplete approximation of f_0 . Beaver et al. (2007) found that some measures of f_0 were marginally significant in their SOF data, so it is possible that f_0 does in fact contribute to the perception of SOF, albeit more subtly than in the case of FOF.

7 Conclusion

7.1 Discussion of results

The first generalization to be tested in this chapter concerned the location of prominence in comparative constructions. According to semantic theory, the location of focus in the comparative clause is determined by the matrix clause. On anaphoric theories of focus, we say that the focus “skeleton” (i.e. a proposition with existential closure over the focused constituent) is related anaphorically to the matrix clause.

In (1a), repeated as (16a), ‘he stayed x long’ entails ‘there is someone who stayed x long’. In (1b), repeated as (16b), the main clause ‘I liked that song x much in the best possible worlds’ entails ‘there are some possible worlds in which I liked that song x much’. In (16c), ‘I understand x much now’ entails ‘I understand x much at some time’.

- (16) a. He stayed longer **than [I]_F did**
antecedent: He stayed x long
- b. I should have liked that song more **than I [did]_F**
antecedent: I should have liked that song x much
- c. I understand even less **than I did** [yesterday]_F
antecedent: I understand even x little

In most cases, the location of prominence can also be predicted according to the (co-) reference of the subjects in the main and comparative clauses (cf. 2). If, as in (16a) the subjects do not co-refer, the subject is focused. If, as in (16b,c), the subjects do co-refer, the focus occurs elsewhere. This provided an objective, text-based criterion by which to classify the data.

The machine learning experiments confirmed the robustness of this generalization. Classifiers trained exclusively on acoustic measurements from web-harvested data achieved accuracy rates as high as 92.9% and balanced error rates as low as 6.5% when tested on similar web-harvested data. They achieved accuracy rates as high as 83.5% and error rates as low as 13.5% when tested on laboratory-elicited data, still well above a baseline of 51.0% accuracy.

Classifiers trained exclusively on acoustic measurements from laboratory-elicited data achieved accuracy rates as high as 90.4% and balanced error rates as high as 9.2% when tested on web-harvested data.

The human classification experiments confirmed the robustness of the generalization, as well. Listeners presented only with web-harvested tokens of ‘than I did’ achieved a mean classification accuracy of 86.4% (standard deviation 8.1%) and a mean balanced error rate of 4.5% (standard deviation 2.8%). Listeners presented only with laboratory-elicited tokens of ‘than I did’ achieved a mean accuracy rate of 78.5% (standard deviation 13.1%) and a mean balanced error rate of 13.1% (standard deviation 6.8%).

From a methodological standpoint, these results demonstrate successful cross-validation of web-harvested and laboratory-elicited data. We might have expected the laboratory-elicited data to allow better human classification and in particular better machine learning classification compared with web-harvested data, since greater variation in recording conditions, discourse context and speaker obtained among the web-harvested data.

Instead, the web-harvested data yielded the better results. Listeners performed better on web-harvested speech than on laboratory-elicited speech; and the performance of web-trained and laboratory-trained classifiers was very competitive when they were tested on web speech. The performance of web-trained classifiers was noticeably poorer when tested on laboratory speech. I attribute these differences to (i) the relative naturalness of the web-harvested data and relatively artificial nature of the laboratory-elicited speech and (ii) the average level of broadcast experience between speakers from the web dataset (especially radio hosts and interviewees) and speakers from the laboratory experiment (viz. university students). Not all of the web-harvested tokens were produced by professional broadcasters, but I would speculate that a high percentage of them have some level of media training or public speaking experience. The laboratory participants, on the other hand, did not have real conversational goals and I speculate that they did not have prior experience being recorded or speaking publicly.

Jennifer Cole (2011) points out that the automatic speech recognition used to index the speech may bias more hyperarticulated speech. Many tokens of speech other than ‘than I did’ were incorrectly transcribed as ‘than I did’; the number and nature of true targets of ‘than I did’ which are incorrectly transcribed as something other than ‘than I did’ is unknown. I suspect that all three factors—the dimensions of naturalness and experience and the performance of the speech recognizer—carry some responsibility for differences observed between the two datasets.

Whatever the source of their variation, both datasets confirmed the correlation between our subject co-reference criterion and the location of prominence in the comparative clause. It should be emphasized here that together the two datasets constitute stronger evidence for this correlation than either alone. A weakness of the web-harvested data is their relative lack of experimental control, which is of course a strength of the laboratory data. A weakness of the laboratory data is their inherent unnaturalness; the naturalness of the web data is one of its strengths.

A second methodological result is the success of the machine learning classifiers, both as a tool for focus detection and as a theoretical model of acoustic realization of prominence and of focus. The classifiers succeeded by the “functional” measure of *mimicing* human behaviour: the classifiers performed on par with or better than humans at the focus classification task. The classifiers also succeeded by the “descriptive” measure of *modeling* to some extent human behavior: the same sets of acoustic features used in the best-performing machine learning classifiers are highly statistically significant predictors of actual listener response.

The particular features selected turned out to have interesting practical and theoretical implications. First, the best *f0* measure ‘mean_f0_ratio’ (the ratio of *f0* means in the vowels of *I* and *did*) was often found *not* to be a good predictor: there was no significant main effect in models of listeners’ judgments and the feature made little or no difference to the performance of the machine learning classifiers. Unfortunately, it is difficult to conclude too much from this apparent non-predictiveness. For example, accurate fundamental frequency measures are

notoriously difficult to extract from a sound file using available algorithms. The classifiers may be improved by more accurate measurements. It is also possible that other kinds of *f0* measurements (e.g. ‘raw’, continuous measures as in Mo 2011) may also provide more robust predictors than the standard measures used here.

The converse finding, however—the predictiveness of non-*f0* measurements in the machine learning classifiers and in the statistical models of listener response—is a significant positive result. The scientific literature on acoustic prominence remains dominated by discussion of fundamental frequency. Kochanski (2006) reported that, in one sample, articles about *f0* outnumbered articles investigating other prosodic cues by nearly 5 to 1. Yet different lines of research have pointed to the robustness of non-*f0* measures. Work in laboratory phonetics and phonology has identified non-*f0* cues of accent in speech production (e.g. Ladefoged 1967, Lehiste 1970, de Jong 1991, Ladefoged and Loeb 2002, Cho 2006) and the acoustics of speech (e.g. Lehiste 1970, Beckman 1986). And Mo (2011) finds that individuals show considerable variation in which combinations of acoustics measures they use to mark prominence and these combinations include *f0* to varying degrees. Finally, we’ve noted that the experimental literature on second occurrence focus has found duration, intensity and amplitude to be small but statistically significant predictors of semantic focus, even when *f0* is not. Phonologists, for their part, have tended to privilege pitch accents over stress. According to “pitch-first” theories (e.g. Bolinger 1958b, Pierrehumbert & Hirschberg 1990, Selkirk 1995a), pitch accent is the primary phonological reflex of focus.

The predictiveness of non-intonational measurements does not refute the role of pitch accent and of *f0* in conveying focus, but it demonstrates that phonological stress and phonetic correlates of stress play a role as well.

According to metrical stress theory, pitch accents can occur only when aligned with a phrase-stressed syllable (e.g. Hayes 1995, Ladd 1980, Liberman 1975, Pierrehumbert 1980, 1993, Selkirk 1984). I find it unparsimonious for focus to be realized by two phonological

categories—pitch accent and stress—and I strongly suspect phrasal stress to be a primary correlate of focus, with pitch accent left to convey additional pragmatic information (e.g. Pierrehumbert & Hirschberg 1990, Steedman 2007), but I leave the matter for future investigation.

Second, the predictive non-*f0* measurements were taken from a single segment, without normalizing to other segments in the utterance: vowel duration and the differential between first and second formants at 40% into the vowel (‘duration_V2’ and ‘f1f2Time40_V2’) came from the vowel of *I* and initial stop closure duration (‘duration_C3’) came from the first consonant in *did*.

It is traditionally held that prosodic prominence is relational or “syntagmatic”, meaning that prominence is processed relative to the sentence that is being uttered (e.g. Jakobson, Fant and Halle 1952; Trubetzkoy 1935, 1939; Lehiste 1970; Ladefoged 1975; Hyman 1978). This explains, among other phenomena, how a word may be perceived as prominent in either fast or slow speech.



(17) than [I]_(F) [did]_(F) *syntagmatic contrast*

Segmental phenomena, such as duration, vowel quality or voice quality, are primarily “paradigmatic”, meaning that they are processed relative to another possible realization. Segmental phenomena, such as the phonological voicing contrast between [p] and [b] are responsible for meaning-distinguishing minimal pairs like *pig* and *big*. There are no minimal pairs in English, so the reasoning goes, that are distinguished solely by intonation (cf. *pig* with a high tone and *pig* with a low tone).

Minimal prosodic pairs (or n-tuples) do exist, however, as we’ve seen (cf. 1-3). According to metrical stress theory (e.g. Liberman 1975, Liberman & Prince 1977; Hayes 1981; Giegerich 1983; Prince 1983; Selkirk 1984; Halle & Vergnaud 1990) prosody is hierarchical, and one can speak of prominence at multiple levels. Prominence at the word level is realized phonologically

by stress, and it is possible to distinguish individual words using stress (e.g. *import* vs. *impórt*). It is only the phrase or utterance level at which prominence is realized intonationally. Thus, one can make intonational contrasts at the phrase or utterance level, but not the word level.

Phonologically, the difference between two minimal intonational pairs is thus both syntagmatic—how prosodic elements are grouped and which prosodic element is most prominent within a grouping—and paradigmatic—how that prosodic structure of one utterance differs minimally from the prosodic structure of another.

Consider the following example from Ladd (1991). He describes an individual “who used to be able to speak German well but then had then spent a long time living in Sweden and now spoke good Swedish but had trouble with German”. Ladd replies to the individual with (18).

(18) **That’s what happened to MY FRENCH.**

It used to be good, but then I spent a year in Germany and ended up with good German, and now whenever I want to speak French I get German interference all over the place.

Semantically, (18) is a case of double focus, on *my* and on *French*. And phonologically, this focus is being conveyed with prominence. Ladd observes, however, that prominence on *my* cannot be purely syntagmatic. It is not the case that *my* is more prominent than its sister, *French*; if anything, *French* is realized with greater prominence than is *my*. The necessary comparison is paradigmatic: (18) is compared to the minimally different realization in (19).

(19) **That’s what happened to my FRENCH.**

Similarly, measures of prominence on *I* alone were good unique predictors in the *than I did* datasets because the salient contrast was not only syntagmatic, but paradigmatic: i.e. between focal and non-focal realizations of *I*.

accented syllable carries a pitch accent and has an unreduced vowel; the prosodic word- accented syllable is postnuclear and has an unreduced vowel; the unaccented syllable has a reduced vowel.

- (22) a. [Was her mama a problem about the wedding?]
 Her **PAPA** posed a problem.
 b. [Did his dad pose a problem as far as their getting married?]
 HER **papa** posed a problem.

This categorical distinction was first proposed by Bolinger (1958,1981) and Vanderslice & Ladefoged (1972) (Gussenhoven 2004:20; see also Halliday 1967). In Table 25, I summarize the properties of the three levels and the examples that illustrate them.

phrase accent	prosodic word accent	accentless
- receives nuclear pitch accent - full, unreduced vowel	- may receive pre-nuclear pitch accent - full, unreduced vowel	- cannot receive a pitch accent - may have some reduced <i>f</i> 0 prominence - reduced vowel
(18) That's what happened to MY FRENCH.		(19) That's what happened to my FRENCH.
(22b) The BUTCHER charged me a thousand bucks	(22a) The butcher charged me a thousand bucks.	
(22a) Her PAPA posed a problem	(22b) HER papa posed a problem	(22b) HER papa posed a problem
than [I] _F did		than I [did] _F [...] _F

Table 25. Summary of categorical prominence contrasts and related examples. Syllable of interest in bold.

Beckman & Edwards observe that the contrast between the prosodic word accented syllable and the unaccented syllable is particularly robust for vowel duration and the degree and speed of jaw opening movement. We can infer that the vowel reduction is also correlated with less extreme formant movement.

In the *thanIdid* datasets, the robustness of measures which are non-intonational and which are extracted only from *I* reflects the categorical and largely paradigmatic prominence on focused *I*. A full, unreduced vowel, as indicated phonetically by greater duration and more extreme formant extrema, is sufficient information to identify the function word as focused with considerably accuracy. It is likely the case that humans use a combination of syntagmatic and

paradigmatic information, and that the choice is context-dependent. In these data, the paradigmatic comparison is particularly salient.

Finally, we are left to interpret the results obtained for the second occurrence focus dataset. According to strong, grammatical accounts of focus association, a focus sensitive construction grammatically requires a semantic focus within the construction. An important empirical test of this is whether the predicted focus is realized prosodically. In (21a), the subject *I* is predicted to be focused, and we've now seen convincing evidence that *I* is realized with prosodic prominence in such examples.

In the second occurrence focus in our dataset, the string *than I did* is repeated: there is a focus in the main clause which has scope (indicated by the Roothian \sim operator) over the entire sentence. In (23b), the focus is on *longer*, since the focus skeleton 'he worked *x* much than I did' is entailed by (23a). We predict *longer* to be realized with greatest prominence in the domain corresponding to (23b), which means that it will have the nuclear pitch accent, and the material that follows (viz. *than I did*) will be subject to pitch reduction.

- (23) a. He worked harder [than [I]_{F1} did] ~ 1
b. and [he worked [longer]_{F3} [than [I]_{F2} did] ~ 2] ~ 3

If the string *than I did* in (23b) is realized without pitch accents, this means that prosodic prominence, if any, must be realized in another way. As we saw in the regular, first occurrence focus data, stress-related measures such as duration and formant extrema are very robust predictors of focus. The grammatical account of focus association predicts that we should observe similarly robust measures on a second occurrence focus if the focus is grammatically required. This prediction is confirmed. The results from the web-based machine learning classifiers initially suggested that the non-intonational measures were somewhat weaker predictors of focus in the SOF data than in the FOF data; the laboratory-trained machine learning classifiers and the human classifiers, however, confirmed the presence of robust non-intonational

predictors: the best laboratory trained classifier achieved 81.0% accuracy (18.1% BER); and although human listener response ranged from 34.4% to 93.8% accuracy (6.1%-34.2% BER), 14 out of the 41 listeners achieved an accuracy rate greater than 81.0%.

Realizing *I* with prosodic word stress—a full, unreduced vowel with more extreme formant extrema and longer duration—is a sufficient cue of prominence and of focus. This observation aligns nicely with the behavior of strong and weak pronouns (e.g. *her* and *'er* in 24-25). Impressionistically, weak or cliticized pronouns are infelicitous or degraded under second occurrence focus (e.g. von Fintel 1994; Rooth 1996a; Krifka 2004; von Fintel attributes this example to Suzanne Tunstall). Strong pronouns have a prosodic word accent; weak pronouns are unaccented (e.g. Selkirk 1996).

(24) Mary's boyfriend only likes HER.
Even her BOSS only likes'er. *weak pronoun*

(25) Mary's boyfriend only likes HER.
Even her BOSS only likes her. *strong pronoun*

Unfortunately, we cannot rule out the possibility that the correlation observed in the SOF data is due to unintended instances of FOF; this would be consistent with the observed variability in the SOF. The poor classification performance by many listeners on the SOF data may also reflect a discrepancy between the experimental task, which asked listeners to make a syntagmatic prominence judgment—which is more prominent: *I* or *did*—and the more salient, paradigmatic task of identifying *I* as accented or unaccented.

7.2 *Future directions*

The success of cross-validating web-harvested and laboratory-elicited data promises considerable benefits. From the perspective of the laboratory experimentalist, web-harvested data provide a way to validate results obtained in lab environments against lingering doubts that speech data collected in the lab do not reflect speech data occurring “in the wild”. The web-

harvested data also provide a rich source of stimuli—both the original, unedited speech and/or transcripts and constructed stimuli informed by the corpus speech. For the corpus linguist, results obtained under the more controlled conditions in lab experiments can inform our analysis of naturally-occurring data, where we find far bigger variability in recording conditions, levels of formality and discourse conditions.

Ideally, the web-harvested data should also be complemented with data from curated speech corpora, provided they are sufficiently large for the phenomenon of interest. Although curated corpora typically represent less diverse speech, unlike web-harvested speech corpora they have complete transcriptions which have been verified. The search space for web-harvested speech is limited to accurately transcribed speech, which introduces a possible bias for more carefully articulated speech.

The hypothesis that focus on function words is realized with local or paradigmatic prominence also requires further investigation. In particular, one can expand the comparatives paradigm to other pronouns (e.g. *than you did*, *than she did*, *than they did*). This also allows one to generalize the classifiers to other vowel types. Formant values may be predictive for all vowel types, but each vowel type will of course exhibit different relationships between formants.

This study also provides a model for other cases where linguistic theory makes predictions for focus location in particular grammatical constructions. Even when theorists agree on perceptual judgments, it is important to test predictions empirically; often there is disagreement over judgments, and experimental examination is the only means of resolving it. The harvest/classification model is particularly useful for constructions which occur with low frequency and are not well represented in curated corpora and/or for which it is difficult to elicit tokens in the laboratory.

Currently, one can search only make string-adjacent searches; it is in part for this reason that the initial study used a pronoun-proverb pair. Eventually, it should be possible to use

wildcards and search for proper names and regular expressions; this will greatly increase the power of the methodology.

CHAPTER 4
FOCUS SENSITIVE OPERATORS BELOW THE PROPOSITION:
THE CASE OF THE ADNOMINAL EMPHATIC REFLEXIVE

1 Introduction

In the next two chapters, I investigate another (allegedly) focus-sensitive construction: the adnominal emphatic reflexive (henceforth adnominal ER), illustrated in (1).

- (1) Jane met **Chomsky himself**.

The adnominal ER exhibits a challenging set of pragmatic, semantic, syntactic and prosodic behaviors, and theoreticians working on it fall into two substantively different (although ultimately reconcilable) camps. In addition to advancing that particular debate, analysis of adnominal ERs also has broader implications for our understanding of focus sensitivity.

In this chapter, I propose that the adnominal ER is a sub-propositional focus operator. Previously, formal semantic architectures for focus, in particular Rooth's (1986, 1992) alternative semantics for focus, have allowed for the possibility of sub-propositional focus operators (i.e. non-truth-conditional constituents). Yet particular semantic accounts of individual focus-sensitive constructions are almost always given in terms of propositional focus operators, leading some theoreticians to conclude that allowing focus-induced interpretations of constituents other than a clause is a specious, but benign, property of the framework (cf. Kadmon 2001:Ch. 18). The arguments for my account of the adnominal ER constitute positive empirical evidence in favor of a sub-propositional focus operator.

Chapter 5 is both an elaboration of these arguments using naturally-occurring data and a case study for another widely held, and rarely challenged, view of focus: that languages can have inherently or obligatorily focused lexical items. I call this property "focal determinism". Both camps of adnominal ER analysts assume that the adnominal ER is always focused and therefore

realized with utterance-level prominence. For one camp, this assumption is a theoretically awkward, idiosyncratic fact of certain focus operators. For the other camp, the perceived prominence is precisely what motivates an account of the adnominal ER as deterministically focused. Chapter 5 is a case study because it is not possible to debunk all alleged instances of focal determinism; the ultimate point is that claims of focal determinism must be tested empirically.

Methodologically, Chapter 4 relies heavily on constructed data and personal introspection (cf. the “theorist’s armchair”). In Chapter 5, I use constructed data to illustrate certain patterns; however, the primary evidence comes from qualitative analysis of individual tokens of naturally-occurring web speech.

This chapter begins by introducing the adnominal ER: what distinguishes it from non-emphatic reflexives and categories of emphatic reflexives, and the two theoretical treatments the adnominal ER has received. In the rest of the chapter, I work through the various properties that, according to the literature, a focus-sensitive operator ought to have, and I review whether or not they hold for the adnominal ER. These include:

- a focus-sensitive operator c-commands its focus associate
- the focus associate of a focus-sensitive operator is realized with prosodic prominence
- a focus-sensitive operator has propositional scope
- a focus-sensitive operator has distinct, focus-sensitive interpretations
- a focus-sensitive operator interacts with other focus operators

I argue variously that these requirements are too restrictive or that they do, in fact, obtain for the adnominal ER. I also consider an analysis of the adnominal ER as focus-insensitive.

Finally, I propose a new analysis of the adnominal ER, and outline the predictions it makes for discourse congruence and prosodic realization, which are tested in Chapter 5. In my analysis, I accept the claim of the first theoretical approach that the adnominal ER is a focus-sensitive

operator. I reject the claim of the other theoretical approach that the adnominal ER is itself invariantly focused; rather, focus on the adnominal ER is just one of several contextually-determined focus configurations.

2 Preliminaries

2.1 What adnominal emphatic reflexives aren't

Emphatic and non-emphatic reflexives

Non-emphatic reflexives always occur as an argument of a verb; emphatic reflexives *never* occur as arguments, but rather as adjuncts. The verb *observe* requires an internal argument and the reflexive in (2) can therefore only be interpreted as this argument.¹ The verb *swim* requires only an external argument; in (3) the external argument is *John*, and so the reflexive can only be interpreted as an adjunct. Prosodically, the non-emphatic reflexive is deaccented: it is realized with less prominence than the verb, in much the same way (perhaps exactly the same) as a non-reflexive pronoun (cf. *What happened to the patient? John observed him.*) The emphatic reflexive is more typically realized with the same prominence as a non-pronominal argument or adjunct (cf. *John swam the channel, John swam vigorously*).²

(2) John observed **himself**.

(3) John swam **himself**.

As further evidence of the distinctness between emphatic and non-emphatic reflexives, consider the co-occurrence of reflexive forms in (4b). Again, since *observe* requires only one internal argument, one of these forms must be interpreted as the internal argument and the other can only be interpreted as an adjunct.

¹ If we allow an intransitive interpretation of *observe* or an interpretation with an elided object, then an adjunct is possible with a meaning similar to *also* or *too*.

² As further evidence, consider the near minimal pair (i-ii).

(i) John served himself.

(ii) John swerved himself.

- (4) a. John observed **himself**.
b. Tom observed **himself himself**.

Agentive ER

Emphatic reflexives further divide into three categories: agentive, additive and adnominal. The agentive emphatic reflexive, illustrated in (5) is distinguished both syntactically and semantically.

- (5) John built the house **himself** last year.

The agentive ER signals the direct involvement of an agent (typically non-assistance or non-delegation). It is most similar in meaning to the *by*-phrase + ER construction, e.g. *by himself*, although it lacks a locational interpretation. Consider the contrast in (6-7). *Too* introduces a presupposition that someone other than Tom doing the dishes alone is salient. In (6), it does not follow from John's washing the dishes himself that he was alone. In (7), however, it does follow from John's washing the dishes by himself that he was alone, and the presupposition is more easily accommodated.

- (6) ? John washed the dishes **himself** yesterday, and Tom did it **alone**, too.
- (7) John washed the dishes **by himself** yesterday, and Tom did it **alone**, too.

Both *himself* and *by himself* have a direct involvement interpretation, from which it is easily accommodated in (8-9) that John washed the dishes without help.

- (8) John washed the dishes **himself** yesterday, and Tom did it **without help**, too.
- (9) John washed the dishes **by himself** yesterday, and Tom did it **without help**, too.

Syntactically, the agentive ER appears to have the distribution of an adjunct, although not an appropriate contrast to the agentive ER is not always clear, making the usual tests for

adjuncthood, such as ellipsis and pseudo-clefting more difficult.^{3 4} In (10a), *himself* is contrasted with a *with*-PP of assistance; in (10b), a factitive construction is contrasted.

- (10) *Ellipsis*
- a. John washed his dishes with Mary's help yesterday; Tom did *(so/it) **himself**.
 - b. My sister had her kids rake up leaves on the weekend, but I did *(so/it) **myself**.
- (11) *Pseudo-clefting*
- a. What Tom did **himself** yesterday was wash the dishes.
 - b. What I did **myself** on the weekend was rake the leaves.

Hole (2002) shows that the agentive ER adjoins to a projection higher than the bare verb phrase (*VoiceP* on his analysis) from their behavior in gerunds.

- (12) Did the neighbors help Grampa rebuild the barn?
- a. *of_{ing}* – nominalization of VP
I remember his rebuilding of the barn (*himself)
 - b. *poss_{ing}* – nominalization of *vP*
I remember his rebuilding the barn (himself)
 - c. *acc_{ing}* – small clause
I remember him rebuilding the barn (himself)

The agentive ER is also sensitive to the aktionsart, occurring only with Vendler-Dowty accomplishment predicates. As indicated by the glosses in (13-16), the felicity of the agentive ER depends on the existence of unique subevents; activities are homogeneous and states either

³ Some tests for the adjuncthood of the agentive ER are even less convincing.

- (iii) *Gapping*
- a. ? John washed the dishes with the help of his wife, and Tom ___ himself.
 - b. ? My sister raked up leaves with her children's assistance, and I ___ myself.
- (iv) *Though parentheses*
- a. ? Wash the dishes though Tom did himself, he didn't enjoy it.
 - b. ? Rake the leaves though I did myself, I didn't enjoy it.

⁴ Culicover & Wilkins (1984) and Adams & Macfarland (1991) have used *himself* as a test for the adjuncthood of other constituents.

- (v)
- a. John wrote the letter himself to Mary.
 - b. John baked the cake himself for Mary.
 - c. *John put the book himself on the table.
 - d. *Mary persuaded Bill herself to leave.
 - e. *Mary promised Bill herself to leave.
- (Culicover & Wilkins 1984)

homogenous or inherently non-eventive and are therefore incompatible with the agentive ER.⁵ Intuitively, if A is the agent of an (exhaustive) event, then she was the agent of all its subevents. Achievements are composed of an activity and a state. Since an activity is homogeneous, and a state non-eventive there is no opportunity for another agent. The examples in (14-16) may be improved, however, if the predicate is understood as substituting for an accomplishment predicate: for example, *found his hotel* in (14a), *got to sleep* in (15a) and *performed an act of kindness* in (16a). This suggests the restriction is pragmatic, rather than semantic.

(13) *Accomplishments*

- a. John built the house himself.
 \approx 'At least one subevent of house building is such that John was the agent'
- b. Mary taught the seminar herself.
 \approx 'At least one subevent of teaching the seminar is such that Mary was the agent'

(14) *Achievements*

- a. # John arrived in Detroit himself.
 \neq 'At least one subevent of arriving in Detroit is such that John was the agent'
- b. # Mary forgot the trombone herself.
 \neq 'At least one subevent of forgetting the trombone is such that Mary was the agent'

(15) *Activities*

- a. # John slept himself.
 \neq 'At least one subevent of sleeping is such that John was the agent'
- b. # Mary waved her hands herself.
 \neq 'At least one subevent of having her hands is such that Mary was the agent'

(16) *States*

- a. #John is a terrific guy himself.
 \neq 'At least one subevent of being a terrific guy is such that John was the agent'
- b. #Mary owns a Porsche herself.
 \neq 'At least one subevent owning a Porsche is such that Mary was the agent.'

⁵ See Tavano (2006) for related observations. She claims that the agentive ER is felicitous with durative predicates, namely accomplishments and achievements, and infelicitous with non-durative predicates, namely activities and states. König & Gast (2006) suggest the contrast is one of dynamicity.

It also follows from the sensitivity to aktionsart that the agentive ER may occur only with external arguments (cf. 17), and therefore cannot occur as the surface subject of object promoting verbs (cf. 17). This also permits a agent-oriented interpretation of event degree modifiers (cf. 18).

- (17) *Object promoting verbs* (cf. Ahn 2009)
- a. #She was drinking herself last night.
 \neq ‘At least one subevent of drinking was such that she was the agent.’
 - b. #The radio broke itself.
 \neq ‘At least one subevent of breaking was such that the radio was the agent.’
 - c. #John arrived himself.
 \neq ‘At least one subevent of arriving was such that John was the agent.’
 - d. #The beef was burned itself.
 \neq ‘At least one subevent of burning was such that the beef was the agent.’
- (18) *Degree modifiers*
- John built the house (mostly / half / partially / all) himself.
 ‘Most / half / part / all of the subevents of building the house were such that John was the agent.’
 \neq ‘John built most / half / part / all of the house’

Additive ER

The additive ER, as the label suggests, is an additive particle, similar to additive focus particle *also*.⁶ Syntactically, it is an adjunct according to tests of ellipsis and pseudo-clefting like the agentive ER and may precede or follow the verbal projection.

- (19) *Ellipsis*
- a. John will wash the dishes, and Tom will (do so) **himself**
 - a'. John will wash the dishes, and Tom will **himself** (do so)
 - b. My sister has raked some leaves, and I will (do so) **myself**
 - b'. My sister has raked some leaves, and I will **myself** (do so)
- (20) *Pseudo-clefting*
- a. What Tom will do **himself** is wash the dishes
 - a'. What Tom will **himself** do is wash the dishes

⁶ Browning (1993) goes so far as to label them ‘*also-reflexives*’.

- b. What I have done **myself** is rake the leaves
- b'. What I have **myself** done is rake the leaves

Like the agentive ER, the additive ER adjoins higher than the bare verb phrase.

- (21) Some of Grampa's neighbors rebuilt their barns, and...
 - a. *of_{ing}* – nominalization of VP
I remember his (*himself) rebuilding of the barn (*himself)
 - b. *poss_{ing}* – nominalization of vP
I remember his (himself) rebuilding the barn (himself)
 - c. *acc_{ing}* – small clause
I remember him (himself) rebuilding the barn (himself)

And as Jackendoff (1972) has noted for other additive particles, the additive ER may follow a first but not a second auxiliary.

- (22) John knows what it means to be honored, because
he is **himself** being (***himself**) honored

Semantically, the additive ER has propositional scope and can associate only with subject arguments (cf. 23-24). The rest of the proposition must be salient of some other individual; the additive ER is therefore infelicitous in “all-new” or “out-of-the-blue” contexts, as in (25).

- (23) John knows what it means to be honored, because
John is being honored **himself**.
- (24) John knows what it means to receive an award from the academy, because
the academy gave John an award (***himself**)
- (25) What's new?
Tom is being honored **himself**.

Unlike the agentive ER, the additive ER may associate with any subject (it is not restricted to external subjects) and is therefore felicitous with object-promoting verbs.

(26) *Object promoting verbs*

- a. Mary was drinking herself last night.
= ‘Mary was drinking last night’
Presupposition: There is someone other than Mary for whom drinking last night is salient.
- b. The radio broke itself.
= ‘The radio broke’
Presupposition: There is something other than the radio for which breaking is salient.
- c. John arrived himself.
= ‘John arrived’
Presupposition: There is some person other than John for whom arriving is salient
- d. The beef was burned itself.
= ‘The beef was burned’
Presupposition: There is something other than beef for which burning is salient.

This can give rise to ambiguities when the additive ER attaches finally: several associations may be possible (cf. 27a). The ambiguity is removed if the additive ER precedes the verbal projection (cf. 27b) or if phi-feature agreement allows only a single association (cf. 28).

- (27) a. Seymour is reputed to have been believed to have been expected to have eaten the bagel **himself**.
(Leskosky 1972:59, cited in Verheijen 1983:283)

- b. Seymour is (**himself**) reputed (**himself**) to have been believed (**himself**) to have been expected (**himself**) to have eaten the bagel (**himself**).

- (28) Gladys believed me to have expected Harry and John to have ordered Seymour to eat the bagel {**herself, myself, themselves, himself**}.
(Leskosky 1972:59, cited in Verheijen 1983:282)

Unlike the agentive ER, the additive ER is felicitous with predicates of all event types (cf. 29-31), but is infelicitous with degree modifiers (cf. 32).

(29) *Achievements*

- a. John arrived in Detroit himself.
‘John arrived in Detroit’
Presupposition: There is someone other than John for whom arriving in Detroit is salient.

- b. Mary forgot the trombone herself.
‘Mary forgot the trombone’
Presupposition: There is someone other than Mary for whom forgetting his/her trombone is salient.
- (30) *Activities*
 - a. John slept himself.
‘John slept’
Presupposition: There is someone other than John for whom sleeping is salient.
 - b. Mary waved her hands herself.
‘Mary waved her hands’
Presupposition: There is someone other than Mary for whom waving his/her hands is salient.
- (31) *States*
 - a. John is a terrific guy himself.
‘John is a terrific guy.’
Presupposition: There is someone other than John for whom being a terrific guy is salient.
 - b. Mary owns a Porsche herself.
‘Mary owns a Porsche’
Presupposition: There is someone other than Mary for whom owning a Porsche is salient.
- (32) *Degree modifiers*
Tom built a house and
John built a house (*mostly/ *half / *partially / *all) **himself**.

Finally, observe that both the adnominal and additive ER may co-occur, although stylistically this is dispreferred.

- (33) Tom will build a house all himself and ...
 - a. John will build a house all **himself himself**.
 - b. John will **himself** build a house all **himself**.
‘For some house x, all subevents of building x are such that John is the agent.’
Presupposition: There is someone other than John for whom building a house himself is salient.

2.2 Syntactic distribution

We are now in a position to identify tests which may distinguish the adnominal ER from the other two categories of ER. We begin with syntactic distribution. With its associated nominal, the adnominal ER may occur in most argument positions, as illustrated in (34).

- (34)
- a. **John himself** will sleep.
 - b. **John himself** has grown up quickly.
 - c. I saw **John himself**.
 - d. I gave **John himself** the torch.
 - e. I gave John **the torch itself**.
 - f. The torch was lit by **John himself**.
 - g. I built the house with **John himself**.
 - h. I persuaded **John himself** to attend the party.
 - i. I promised **John himself** to attend the party.
 - j. I remember **John himself** building the house.

Degree modification can be used as a test to distinguish the adnominal ER from the agentive ER: only the agentive ER is compatible with degree modification.

- (35) John (#mostly/#partly) himself has built this house.

And for non-subject associates, the non-equivalence of *also* is a convenient test for distinguishing the adnominal ER from the additive ER (cf. *I saw John himself* ≠ *I saw John also*).

- (36) I saw John himself
 ≠ ‘I saw John also’

The distinction between additive ERs and adnominal ERs in subject position is more subtle and will be taken up again in more detail when discussing prosody in Chapter 5. Consider, however, the distribution of postnominal DP-modifying *only* and pre-auxiliary VP-modifying *only*. In examples without overt auxiliaries like (37), the two are string-ambiguous: association

is possible within the nominal or verbal projection. The pre- or post-auxiliary position in the string clearly disambiguates the two readings (cf. 38-39 and 40-41).⁷

- (37) The President **only** persuaded Congress.
 - a. = ‘No one other than the President persuaded Congress’
 - b. = ‘The President persuaded no individuals other than Congress’
- (38) The President **only can** persuade Congress.
 - a. = ‘No one other than the President can persuade Congress’
 - b. ≠ ‘The President can persuade no individuals other than Congress’
- (39) The President **can only** persuade Congress.
 - a. ≠ ‘No one other than the President can persuade Congress’
 - b. = ‘The President can persuade no individuals other than Congress’
- (40) The Provost **only has** been successful at raising funds.
 - a. = ‘No one other than the Provost has been successful at raising funds’
 - b. ≠ ‘The Provost has been successful at nothing other than raising funds’
- (41) The Provost **has only** been successful at raising funds.
 - a. ≠ ‘No one other than the Provost has been successful at raising funds’
 - b. = ‘The Provost has been successful at nothing other than raising funds’

A possible exception to the generalization that [XP *himself*] may occur in all different argument positions is the Saxon genitive ‘s.

- (42) a. ?I met **John himself**’s daughter.
- b. I remember the daughter of John **himself**.
- c. ? I remember John **himself**’s building of the house.
- d. I remember the building of the house by John **himself**.

This restriction is claimed to apply equally for the subject of a gerund participial (Verheijen 1983:259-60; Huddleston & Pullum 2002:1497).

- (43) a. They objected to Tom/Tom’s doing it.
- b. They objected to ?Tom himself’s doing it.

⁷ As discussed in Chapter 4, the auxiliary position diagnostic is more complicated for an ER since a focused adnominal ER may result in a (slightly more restrictive) additive interpretation as well.

The robustness of this exception is unclear. One can find many attested written examples of the adnominal ER with the genitive, both in edited volumes and on the web. Examples (44-49) contain instances of the adnominal ER in a simple genitive.

- (44) *I got 'em, and Granny said you'd like to see them, so she did--and here's what will please you--see my certificates--see, signed by **the doctor himself's** own hand and Father M'Cormuck, that's his name, with his blessing by the same token he gave me.*
from Chapter 24 of *Ormond* (1817) by Maria Edgeworth
- (45) *I am able to stress the several aspects of Melville's thought on this because, note, in each case the feeling or necessity of the inert, or of passivity as a position of rest, is joined to the most instant and powerful actions Melville can invent: **the whale itself's** swiftness, Ahab's inordinate will, and the harpooneer's ability to kill from calm only.*
In *Selected Writings* by Charles Olson, pp. 51-52 (1997)
- (46) *Where Marius claimed an acquaintance with Septimius' niece, Claudia's father was **the emperor himself's** dearest friend from university days in Rome.*
from *In Perpetua: A Bride, a Martyr, a Passion* (2004) by Amy Rachel Peterson, p. 291
- (47) *With far more content than Huw's rivals, Team Lewis has clearly been preparing their ground. Bearing in mind his Diary feature starts in June 2007 (when he left the Welsh Assembly Government) and seems to incorporate his own blog and the 20:20 feed, this has been a project a long time in the making.*
Much like **the man himself's** leadership campaign...
[web example]
- (48) *Ronson himself takes lead vocals on Lose It (In The End)*
he took singing lessons from Lady Gaga's vocal coach, at **the Lady herself's** suggestion
[web example]
- (49) *If you can play this, you're just plain awesome.*
Details, such as **the song itself's** name, are in-video.
[web example]

One can also find naturally-occurring examples of adnominal ER in the subject of gerunds, as in (50-52).

- (50) *No matter how sweetly that invitation were made, this would still, finally, be much less than **the LORD himself's having** come down from heaven in the form of a human being [...]*
From *Treasury of C. F. W. Walther - Volume II: Festival Sermons and Prayers from Lent through the Easter Season* (2008), translated by Joel Baseeley, p. 40.
- (51) [...] this cannot take place without **the soul itself's becoming** a prey to the passions of, for example, grief or anger.
Plotinus: the road to reality, by John M. Rist 1967, p. 176
- (52) [...] for when Bach "makes the charts" without **the music itself's having** been tampered with, there is little doubt that Bach himself deserves the credit.
American Record Guide: Volume 30, 1963

Grammaticality judgments of postnominal *only* and *alone* in the same utterances are more robust (cf. 53-54), and I cannot find any naturally-occurring examples.

- (53) * Claudia's father was **the emperor only/alone** 's dearest friend from university days in Rome.
Intended: 'Claudia's father was the dearest friend of only the emperor.'
- (54) * The **song only/alone** 's name is available in-video.
Intended: 'The name of only the song is available in-video.'

In section 3.3, I attribute this difference between *himself* and *only/alone* to syntactic scope at logical form. *Only* and *alone* have propositional scope; they and their DP complements are subject to constraints on movement; *himself* has DP-scope and is not required not move.

2.3 Constituency

The adnominal ER is also distinct in forming a constituent with its associate. Siemund (2000) applies a battery of constituency tests to establish their syntactic relationship, including the following. The postnominal focus operators *only* and *alone* show the same syntactic behavior.

Stand-alone

- (55) a. Who mowed the lawn? **John himself/only/alone**.
b. What did he mow? The lawn **itself/only/alone**.

Object NP deletion

- (56) a. The children ate **the cake itself/only/alone**
b. The children ate ____ (***itself/*only/*alone**)

Pronoun replacement

- (57) a. **The director himself/only/alone**, who ...
b. **Paul himself/only/alone** thinks that he ...

Clefting

- (58) It was **the faulty switch itself/only/alone** that ____ caused the trouble.

Pseudo-clefting

- (59) a. **The faulty switch itself/only/alone** caused the trouble.
b. What caused the trouble was **the faulty switch itself/only/alone**.

Topicalization

- (60) **Ben Nevis itself/only/alone**, I would like to climb ____.

Inchoatives

- (61) a. Janet broke **the cup itself/only/alone**.
b. **The cup itself/only/alone** broke ____.

Dative alternation

- (62) a. Bill sent Tom **the book itself/only/alone**.
b. Bill sent **the book itself/only/alone** to Tom.

2.4 Agreement and c-command

Like all reflexives, the adnominal ER must agree in phi-features, namely person, number and gender, with its associate. It is this requirement which is responsible for the co-occurrence restrictions listed in (63).

- (63) a. the cake **itself** / ***himself** / ***herself** / ***themselves** / ***ourselves**
b. the boys ***itself** / ***himself** / ***herself** / **themselves** / ***ourselves**

We might also reasonably assume that the agreement is syntactically licensed by c-command of the adnominal ER by its associate (cf. Principle A of standard binding theory).⁸

⁸ Although see discussion of additive ER. Unlike the adnominal ER, it is hard to see how the additive ER and its associate could mutually c-command each other.

Other postnominal focus operators in English lack agreement morphology; indeed focus operators in Germanic are notoriously uninflected (cf. Siemund 2000:14). One need only look as far as French⁹, however, to observe agreement in a postnominal focus marker similar to English *alone*. The French form *seul*, which like *alone* may also occur as a predicative adjective or verbal modifier, must agree in number and gender with the associating nominal¹⁰.

- | | | |
|------|--|-----------------------------|
| (64) | a. Lui seul/*seule/*seuls/*seules sait pourquoi. | ‘He alone knows why.’ |
| | b. Elle *seul/seule/*seuls/*seules sait pourquoi. | ‘She alone knows why.’ |
| | c. Eux *seul/*seule/seuls/*seules savent pourquoi. | ‘They (m.) alone know why.’ |
| | d. Elles *seul/*seule/*seuls/seules savent pourquoi. | ‘They (f.) alone know why.’ |

2.5 Previous formal semantic accounts

Virtually all formal semantic accounts of adnominal ERs maintain that the meaning of the adnominal ER is sensitive to focus.¹¹ How this sensitivity to focus is achieved divides the literature into two general approaches: the FOCUS SENSITIVE OPERATOR approach (e.g. König 1991, Siemund 2000, Bergeton 2004) and the FOCUSED ASSERTION OF IDENTITY approach (e.g. Eckardt 2001, Hole 2002, Gast 2006).

⁹ König & Gast (2006) cite Old English *self/seolf/sylf* which inflects for gender, number and case.

¹⁰ Focused pronouns in French always occur in their oblique form (called “pronoms disjoints” in traditional grammars). In the absence of *seul*, (64) would be realized as in (iii).

- | | | |
|------|---------------------------|-----------------------|
| (vi) | a. Il sait pourquoi. | ‘He knows why.’ |
| | b. Elle sait pourquoi. | ‘She knows why.’ |
| | c. Ils savent pourquoi. | ‘They (m.) know why.’ |
| | d. Elles savent pourquoi. | ‘They (f.) know why.’ |

It is this form of the pronoun which is used with other uncontroversial focus operators like *aussi* ‘too’.

- | | | |
|-------|--|--|
| (vii) | a. Il sait pourquoi, lui aussi. | ‘He knows why, too’ [<i>lit.</i> He knows why, him also.] |
| | b. Elle sait pourquoi, elle aussi. | ‘She knows why, too.’ [<i>lit.</i> She knows why, her also.] |
| | c. Ils savent pourquoi, eux aussi. | ‘They (m.) know why, too.’ [<i>lit.</i> They (m.) know why, them (m.) also] |
| | d. Elles savent pourquoi, elles aussi. | ‘They (f.) know why, too.’ [<i>lit.</i> They (f.) know why, them (f.) also] |

¹¹ See Cunningham (2009) for arguments against a focus-based account and in favor of a scalar account. I discuss Cunningham’s analysis in Section 3.4.2.

According to the FOCUS SENSITIVE OPERATOR approach, *himself* is an operator which requires a focused argument, e.g. *the Provost*. The adnominal ER is one of a handful of adnominal focus-sensitive operators in English, including *only* and *alone*, which follow their focused nominal argument (cf. *The Provost alone will chair the committee*, *The Provost only will chair the committee*). We can annotate sentence (1) in the framework of Rooth (1992): the focus interpretation operator \sim fixes the scope of *himself* and syntactic F-marking fixes the focus associate. The set of alternatives is context-sensitive but focus constrains the members of the set to those with a nominal form (e.g. individuals with type $\langle e \rangle$).

- (65) FOCUS SENSITIVE OPERATOR
 [[[the Provost]_F \sim] *himself*] will chair the committee

Alternative set: the set of individuals $\langle e \rangle$
 e.g. {the President, the VP-Academic, the Dean of Science...}

The FOCUSED ASSERTION OF IDENTITY approach (e.g. Eckardt 2001, Hole 2002, Gast 2006) maintains that *himself* is not a focus-sensitive operator. Semantically, it is syncategorematic, i.e. the identity operator, which returns the value of whatever nominal argument is supplied to it. It is not the argument of this function which is focused; rather, it is *himself* which is focused. Indeed, on this approach, the adnominal ER is invariably focused, and it is only in this capacity that *himself* contributes meaningfully to the interpretation of the sentence. In the notation of Rooth (1992), then, *himself* is syntactically marked for focus and the focus interpretation operator must occur higher syntactically higher than *himself*, typically at the clausal level. Alternatives to the identity function are other contextually-salient relational predicates, e.g. functions $\langle ee \rangle$ from individuals to other individuals.

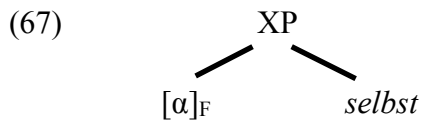
- (66) FOCUSED ASSERTION OF IDENTITY
 [the Provost [*himself*]_F will chair the committee] \sim]
 Alternative set: { ASSISTANT-TO(the Provost), NOMINEE-OF(the Provost), WIFE-OF(the Provost) ... }

Deciding between the two approaches amounts to more than a technical, formal exercise. Critics of the FOCUS SENSITIVE OPERATOR approach argue that the adnominal ER simply does not behave like a focus sensitive operator should, raising an important theoretical question: what *are* the necessary and sufficient properties of a focus-sensitive operator? In the following section, I compare the behavior of the adnominal ER against requirements that have been proposed in the literature for focus sensitivity.

3 *Properties of focus sensitive operators*

3.1 *A focus sensitive operator c-commands its associate*

Eckardt (2001) presents a detailed analysis of stressed *selbst* on the FOCUSED ASSERTION OF IDENTITY approach. Against the FOCUS-SENSITIVE OPERATOR approach, she asserts that the putative focus operator does not c-command its putative associate. While this is a plausible assertion for occurrences of *selbst* which like the additive and agentive ER in English adjoin to a verbal projection and do not c-command an external argument, it remains to be shown that adnominal *selbst* does not c-command its associate.



I will not make any claims about the German data; however, it is no longer clear that c-command is even a prerequisite for all cases of association with focus in English or German. In a survey of a variety of focus sensitive constructions, Beaver & Clark (2008) observe that while many focus sensitive constructions are lexically specified for focus (conventional association), and require an operator to c-command its associate, many focus sensitive constructions do not.

Conventional association	Free association
Exclusives	Quantificational adverbs
Scalar additives	Quantificational determiners
Non-scalar additives	Generics
Particularizers	Counterfactuals
Minimizing downtoners	Reason clauses
Maximizing downtoners	Emotive factives
Intensifiers	Verbs of desire
	Modals

Table 1. Partial classification of focus sensitive expressions modified from Beaver & Clark (2008:78)

The exclusive particle *only*¹², for example, must c-command its associate, while the quantificational adverb *always*¹³ often does not. This behavior is shown in particular by extraction of the putative focus associate¹⁴ from a variety of different contexts.

In (68), *always* may associate with *what*, yielding the reading in (a), or *his mother*, yielding the reading in (b). In (69) by contrast, *only* cannot associate with *what* with the reading in (a), although it can associate with *his mother* to give the (b) reading. According to Beaver & Clark, an operator like *only* must associate with a lexical constituent in its c-command domain¹⁵. (69) is ungrammatical on the (a) reading because the associate *what* has been extracted outside of this c-command domain. The same reading is available for *always* because it is not subject to this same constraint.

Wh-questions

- (68) What do you think Karl **always** gives his mother ____ ?
a. ✓ Association with *what*
‘What is the thing such that Karl gives that thing and nothing else to his mother?’

¹² Beaver & Clark (2008) test only VP-*only*.

¹³ According to Beaver & Clark, focus constrains the domain variable introduced by *always*.

¹⁴ The authors also use VP-ellipsis as a test. Unfortunately this is not a useful test for the post-nominal focus operators.

¹⁵ They cite this principle of lexical association in Aoun and Li 1993:206, due to Tancredi (1990).

- b. ✓ Association with *his mother*
‘What do you think gives his mother and no one else?’
- (69) What do you think Karl **only** gives his mother ____ ?
- a. ✗ Association with *what*
‘What is the thing such that Karl gives that thing and nothing else to his mother?’
 - b. ✓ Association with *his mother*
‘What do you think gives his mother and no one else?’

Beaver & Clark observe the same pattern for other extraction contexts, including *wh*-relatives, topicalization, adverb preposing and inverted *wh*-clefts: *only* cannot associate with a constituent that has been moved outside of its c-command domain; the same association for *always* is perfectly grammatical. (Asterisks in the following examples indicate ungrammaticality for association with the extracted constituent.)

Wh-relatives (cf. Krifka 1992)

- (70) We should thank the man whom Mary **always** took ____ to the movies.
- (71) *We should thank the man whom Mary **only** took ____ to the movies.

Inverted wh-Cleft

- (72) Guinness is what I think Kim **always** wants to drink ____
- (73) *Guinness is what I think Kim **only** wants to drink ____

Topicalization

- (74) Fishsticks, I believe Kim **always** buys ____
- (75) *Fishsticks, I believe Kim **only** buys ____

Adverb Preposing

- (76) On Sunday, I thought you **always** went to the store ____
- (77) *On Sunday, I thought you **only** went to the store ____

Given this taxonomy of focus sensitive constructions, it does not follow that lack of c-command over a putative associate excludes a focus-sensitive operator account of a given particle. Nonetheless, these same tests do indeed support a conventionally focus-sensitive operator account of the adnominal ER, as well as other post-nominal focus operators *only* and *alone*.

Consider again the *wh*-question context. In (78), association with *his mother* is ungrammatical it does not agree in animacy with *itself*; the desired association with *what* is ungrammatical, by hypothesis, because it has been extracted outside of its c-command domain. It is also for this reason that post-nominal *only* and *alone* cannot associate with *what*.

Wh-questions

- (78) What do you think Karl gives his mother ____ **itself**?
- a. ✗ Association with *what*
‘What is the thing such that Karl gives that thing and that thing is really likely/important/salient/etc.?’
 - b. ✗ Association with *his mother*
‘What do you think Karl gives his mother and his mother is really likely/important/salient/etc.?’
- (79) What do you think Karl gives his mother ____ **only/alone**?
- a. ✗ Association with *what*
‘What is the thing such that Karl gives that thing and nothing else to his mother?’
 - b. ✓ Association with *his mother*
‘What do you think gives his mother and no one else?’

One may suspect that the ungrammaticality of (79) is due simply to a restriction on the adnominal ER with quantifiers more generally. However, association is equally ungrammatical with extraposed constituents that are not *wh* expressions (84-89). Again, the asterisks indicate ungrammaticality for association with the extracted constituent.

Wh-relatives

- (80) * We should thank the man whom Mary took ____ **himself** to the movies.
- (81) * We should thank the man whom Mary took ____ **only/alone** to the movies.

Inverted Wh-Cleft

- (82) *Guinness is what I think Kim wants to drink ____ **itself**
- (83) *Guinness is what I think Kim wants to drink ____ **only/alone**

Topicalization

- (84) *Fishsticks, I believe Kim buys ____ **themselves**
- (85) *Fishsticks, I believe Kim buys ____ **only/alone**

Adverb Preposing

- (86) *On Sunday, I thought you went to the store ____ **itself**
(87) *On Sunday, I thought you went to the store ____ **only/alone**

That-relatives

- (88) *There is one problem that the students couldn't solve ____ **itself**
(89) *There is one problem that the students couldn't solve ____ **only/alone**

Siemund (2000) observes a number of other ungrammatical ellipsis and extraction contexts.¹⁶ Post-nominal *only* and *alone* pattern similarly.¹⁷

Object NP deletion

- (90) a. Mike ate the cake **itself**
b. Mike ate (***itself**)

(91) a. Mike ate the cake **only/alone**
b. Mike ate (***only/*alone**)

Passivization

- (92) a. The bottle **itself** was fetched ____
b. The bottle was fetched ____ **itself**

(93) a. The bottle **only/alone** was fetched ____
b. The bottle was fetched ____ **only/alone**

Clefting

- (94) a. The faulty switch **itself** caused the trouble.
b. It was the faulty switch **itself** that ____ caused the trouble.
c. *It was the faulty switch that ____ **itself** caused the trouble.

(95) a. The faulty switch **only/alone** caused the trouble.
b. It was the faulty switch **only/alone** that ____ caused the trouble.
c. *It was the faulty switch that ____ **only/alone** caused the trouble.

¹⁶ Siemund takes these data only as evidence of co-constituency of the adnominal ER and its associate.

¹⁷ Note that we are interested in post-nominal reading of *alone*. In many cases, there is an acceptable reading of these sentences with *alone* indicating lack of co-agency or whether or not some other individual was present.

Inchoatives

- (96) a. Janet broke the cup **itself**.
b. The cup **itself** broke.
c. *The cup broke **itself**.
- (97) a. Janet broke the cup **only/alone**.
b. The cup **only/alone** broke.
c. *The cup broke **only/alone**.

Dative alternation

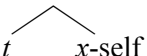
- (98) a. Bill sent Tom the book **itself**.
b. Bill sent the book **itself** to Tom.
c. *Bill sent the book to Tom **itself**.
- (99) a. Bill sent Tom the book **only/alone**.
b. Bill sent the book **only/alone** to Tom.
c. *Bill sent the book to Tom **only/alone**.

All of these tests are consistent with the principle of lexical association. Since VP-*only* is the paradigmatic case of lexical association with focus, it is highly desirable for post-nominal *only* to follow this principal also.¹⁸

We concluded from Siemund's constituency tests discussed in the previous section that the adnominal ER forms a constituent with its associate. The extraction tests we've just reviewed suggest that the adnominal ER also c-commands its associate. If the agreement between the adnominal ER and its associate must be licenced by c-command of the associate over the adnominal ER, then this means that the adnominal ER stand in a relationship of mutual c-command (i.e. sisterhood).

Siemund (2000) argues that the adnominal ER attaches above other adjuncts, including relative clauses, although some of ungrammaticality examples (viz. 101,102) are also consistent

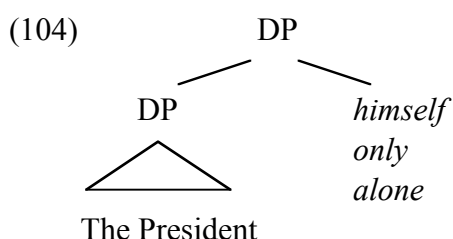
¹⁸ Alternatively, the ungrammaticality of the extraction contexts may be due to a structural prohibition, e.g. an island constraint, on a branching configuration with a trace.

(viii) 

with a constraint on extraction out of complex NPs. Post-nominal *only* and *alone* follow the same pattern.

- (100) a. the room [underneath the kitchen] **itself/only/alone**
 b. ?? the room itself [underneath the kitchen]
- (101) a. Henry [the Sixth King of England] himself/only/alone
 b. ?? Henry **himself/only/alone** [the Sixth King of England]
- (102) a. the man [that you met] **himself/only/alone**
 b. ?? the man **himself/only/alone** [that you met]
- (103) a. the belief [that linguistics is easy] **itself/only/alone**
 b. ?? the belief **itself/only/alone** [that linguistics is easy]

This leads us to a syntactic representation like (104).



3.2 *A focus associate has a focus exponent*

The next property assumes the reasonable, although not uncontroversial premise, that semantic focus corresponds to phonological prominence. I start with what Ladd (1991) terms the “commonsense view” of prominence common in the semantic and syntactic literature on focus: “something is stressed or it isn’t”.

Analysts of the adnominal ER, from both camps, take for granted that the adnominal ER *isn’t* stressed. Advocates of the FOCUS-SENSITIVE OPERATOR approach (e.g. König 1991, König & Siemund 1996) admits the lack of prominence as an idiosyncratic fact of the adnominal ER and of postposed focus operators more generally; critics of the approach (e.g. Eckhardt 2000, Hole 2002; König & Gast 2006) see the lack of prominence as a major failing. (Eckhardt and Hole’s

analyses are meant to account for German *selbst*, although I take it their application to English adnominal ER is implied. I make no claims about German here.)

In Sections 3.2.1 and 3.2.2, I review evidence that the associate can be and often is stressed, on the commonsense view. Associates in Ahn's (2009) study of prosody in emphatic reflexive constructions were frequently labeled with pitch accents by trained annotators. And in tokens of the adnominal ER harvested from the web, more than half of the associates were measureably higher *f0* maxima than the ER.

In Section 3.2.3, I consider a more nuanced, abstract view of prominence from phonological theory, according which prominence involves structured relations between prosodic units within a particular domain and in which the prosodic domain corresponds to the scope of focus. On this view, it is possible for the associate of the adnominal ER to be prominent within one domain, and non-prominent within another.

3.2.1 *Qualitative evidence*

Ahn (2008) is the only previous study on the prosody of emphatic reflexives that I am aware of. Three native English speakers read scripts containing emphatic reflexives, 24 scripts in total, 8 of which contained "adjacent emphatics". Ahn defines this class of emphatic reflexives syntactically by its adjacency to the nominal; they therefore belong to what I am calling adnominal ER.¹⁹

The utterances were labeled prosodically by two trained annotators using the MAE_ToBI conventions (cf. Beckman *et al.* 2005). In this framework, intonational prominence is largely categorical, syllables either having or lacking a pitch accent. The framework can also capture many important subtleties such as different accent types and downstep; however the model is

¹⁹ Semantically, Ahn describes this class as having an additive interpretation uniquely, something I reject for AER below, and doubtless the scripts were designed with this analysis in mind. If my analysis below is correct, then Ahn's data are biased towards a particular prosodic realization--namely greatest prominence on the adnominal ER itself--although we'll see that this bias does not undermine the claim in this section that the nominal does not lack prominence. If anything, this bias should disfavor prominence on the associated nominal.

meant to capture categorical phonological generalizations, rather than fine-grained, gradient phonetic distinctions. On the hypothesis that the associated nominal of an adnominal ER lacks prominence or is deaccented, we should expect the nominal to lack a pitch accent. Instead, Ahn found that among all 122 occurrence of the emphatic reflexive (including “adjacent emphatics”, “post-VP emphatics” and “sentence final emphatics”), the nominal was labeled with a pitch accent in approximately 75% of occurrences.

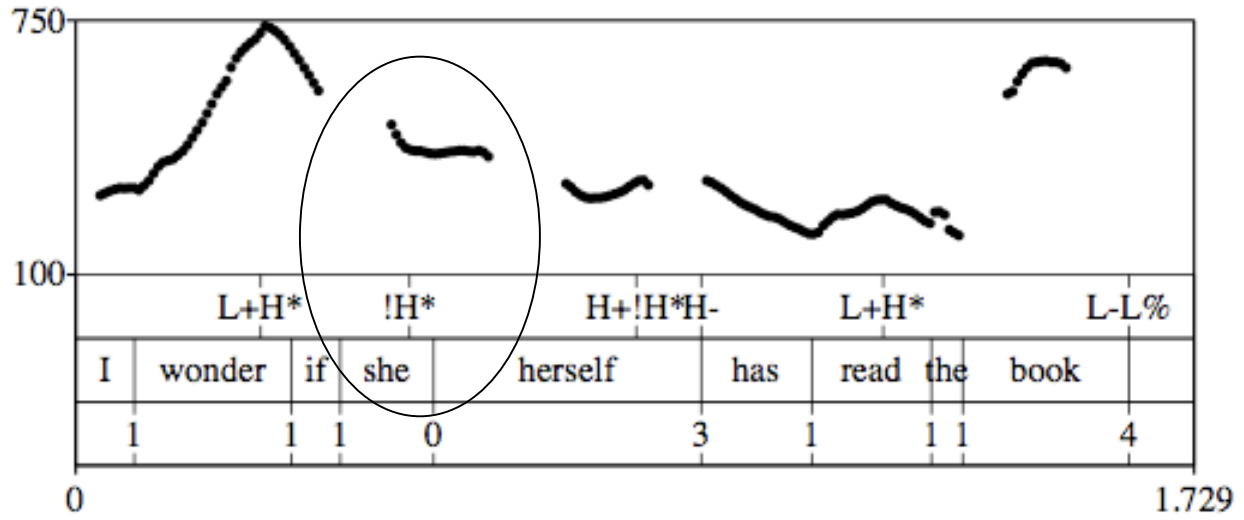
Ahn reports that in approximately 60% the utterances containing an adnominal ER, the associated nominal was labeled with a high pitch accent (H*). (He does not report in how many cases the nominal was labeled with some other pitch accent type.) He does include 10 unique pitch tracks and corresponding MAE_ToBI annotations of sentences containing adnominal ERs. The nominal is labeled with a pitch accent in 9 out of these 10 reported cases.

Below are the annotations of utterances from the same written script by three different participants. Although variation exists elsewhere in the utterance, the pronoun *she* is realized with a pitch accent by all three speakers. [The H* symbol represents a simplex high pitch accent.]

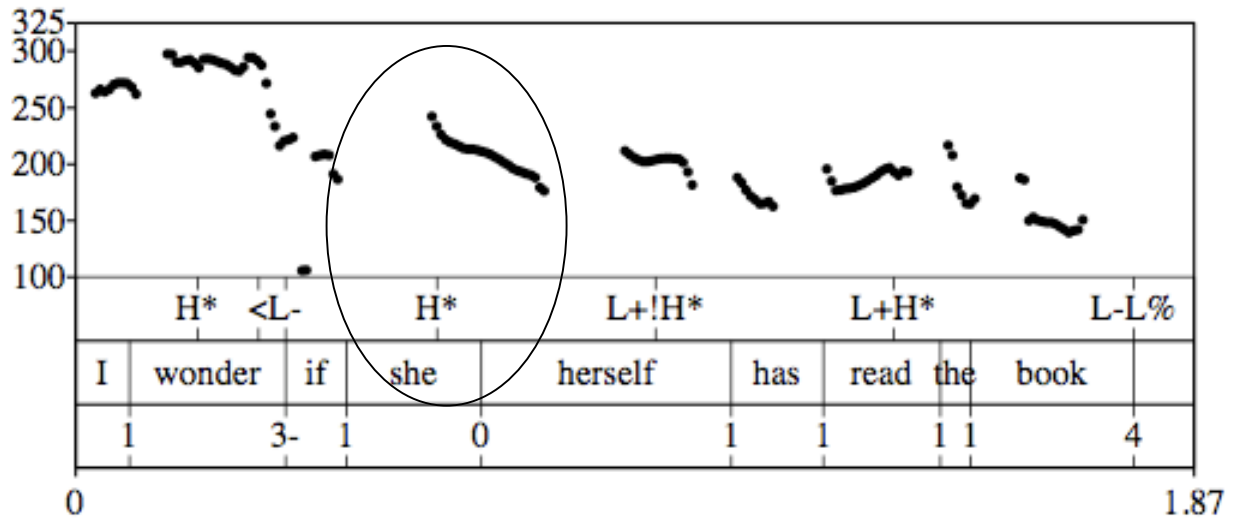
Not only were many of the nominal associates in Ahn’s data realized with a H* pitch accent, the majority had pronominal associates: 5 out of the 8 adnominal ER stimuli. This is significant since, as noted in Chapter 3, function words like pronouns tend to lack intonational and even non-intonational prominence (i.e. pitch accent and phrasal or word stress), unless focused.

The pronominal data are thus strong evidence that the associated nominal of an adnominal ER are realized with prosodic prominence. Given only a categorical notion of prominence—prominent or not prominent—we can say the nominal is frequently realized with prominence.

BW, Script 7-2



KV, Script 7-1



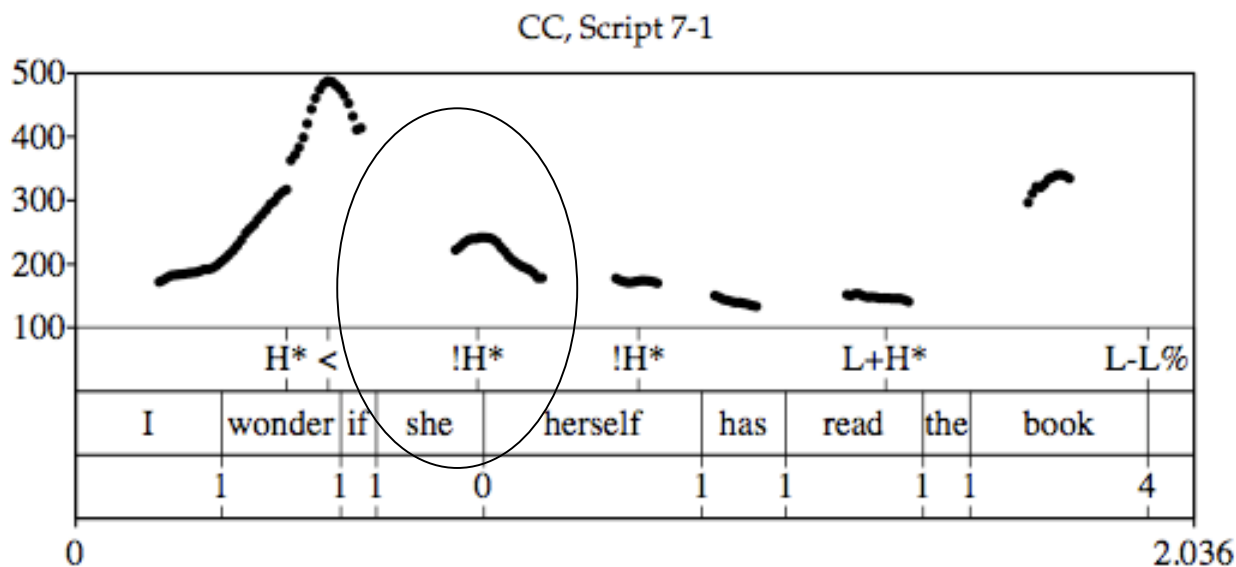


Figure 1: *F0* tracks and MAE_ToBI annotations from Ahn (2008). I have added circles around the associate of the adnominal ER.

3.2.2 Quantitative evidence

As described in Chapter 2, I harvested a small corpus of naturally occurring tokens for the adnominal ER. From an initial 300 purported tokens of the target *he himself*, 232 were unique, true tokens. While there is no simple, objective measure for the presence of pitch accent, as a crude measure of relative phonetic prominence, one can compare the value of maximum *f0* in the interval corresponding to the nominal and the value of the maximum *f0* corresponding to the stressed syllable –*self* of the adnominal ER. On the hypothesis that the associated nominal lacks prominence, we expect the *f0* maximum in the –*self* interval to be consistently greater than the *f0* maximum in the *he* interval.

In 127 of the 232 utterances in the web-harvested corpus (approximately 55%), the maximum *f0* for *he* was greater than the maximum *f0* for –*self*. By this measure, then, in roughly half of the utterances, the associated nominal was indeed realized with intonational prominence.

3.2.3 Prominence in a domain

While prosodic prominence can sometimes be paradigmatic, as it is on the “commonsense” view (cf. Ladd 1991), phonologists have noted for many years that prominence is also relational

(Fox 2000 cites Jakobson 1931/1962 and Trubetzkoy 1935, 1939 as early modern examples). For example, intonational prominence is one sense absolute: either the syllable has pitch accent or it does not. Given two or more pitch accents in a phrase, however, we can speak of prominence relative to that domain. Similarly, an *f0* value (or other measure) can be perceived as an extremum only relative to some other value or set of values. In the previous subsection, I compared local maxima within the *he* interval and *himself* interval. By this phonetic measure, *he* was frequently prominent within the domain corresponding to *he himself*.

Many linguists investigating focus and prominence have proposed mapping the semantic scope of focus to a phonological domain (e.g. Chomsky 1971, Jackendoff 1972, Truckenbrodt 1995, Zubizarreta 1998, Büring 2008). The formalism in (105) is adapted from Rooth (2009).

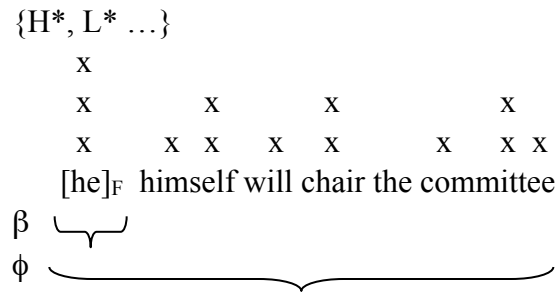
(105) *Stress-F* (based on Rooth 2009)

Let β be an F-marked phrase with scope ϕ . Then the most prominent syllable in the phonological realization of ϕ falls within the realization of β .

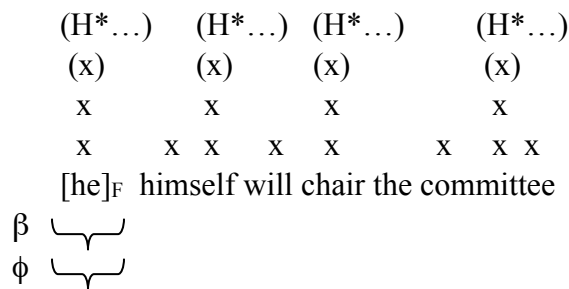
Let us apply this principle to the FOCUS-SENSITIVE OPERATOR approach to adnominal ERs. The associate (e.g. *he*) will be the F-marked (focused) phrase β . First, let us suppose that the focus has scope ϕ over the entire clause; in that case the associate must be realized with greatest prominence within a domain β corresponding to the clause. Other syllables of the clause will be relatively less prominent (cf. 106a). It is this prosody which both approaches assume does not occur.

According to the proposals of König (1991) and Siemund (2000), the scope of the focus ϕ is not clausal. The scope ϕ and focus β of the adnominal ER are in fact co-extensive, meaning that the most prominent syllable within the realization of the associate will necessarily be contained within the same domain: the DP. The analysis has nothing to predict about the prosody of other material in the clause. In (106b), I represent this by marking stress and pitch accent with parentheses as optional.

(106) a. *Clausal scope*



b. *DP-scope*



There is general agreement among phonologists that prosodic prominence is also hierarchical, with relative prominence existing in English at the word level (e.g. Chomsky & Halle 1968), at the utterance level, and other levels in between (e.g. Selkirk 1980, Nespor & Vogel 1982, Beckman & Pierrehumbert 1986). Phonologists also hold that intonational events (i.e. tones or pitch accents) belong to a semi-autonomous, highest level of prominence (Lieberman 1975; Pierrehumbert 1980). Each level of prominence, in turn, corresponds to a different prosodic unit.

The existence of different prosodic units means that the scope of focus need not be mapped to the same prosodic constituent in all utterances; similarly, prominence on the focus need not be mapped to the same level of prosodic prominence. Typically, the sentence corresponds to the intonational phrase and greatest prominence within that domain will be the last or “nuclear” pitch accent. The focus in (106a) is predicted to be realized with a nuclear pitch accent; potential pitch accents occurring after the nuclear pitch accent are suppressed. In (106b), by contrast, the associate *nav* be realized with a nuclear accent: but other levels of prominence will also yield

satisfy Stress-F. Indeed, experimental studies of second occurrence focus (e.g. Rooth 1996b and Beaver *et al.* 2007; cf. Chapter 3) have shown that a putative focus associate may lack a pitch accent entirely in its prosodic realization, but may nonetheless be realized with greatest prominence (e.g. duration, intensity) within the domain corresponding to the scope of focus.

It may be the case that there is a minimal degree of prominence entailed by focus marking. Even in those cases where a focus associate is realized without any pitch accents, it is claimed that the constituent must be realized at least as a syllable or prosodic foot. For example, a pronoun may not be reduced or cliticized when focused, even if it lacks intonational prominence, as the repeated pronoun in (107) does.²⁰

- (107) a. Mary's boyfriend only likes HER.
 Even her BOSS only **likes her**.
 b. Mary's boyfriend only likes HER.
 # Even her BOSS only **likes'er**.

Applying the same test to the associate of an adnominal ER, I find the reduction of *he* to 'e infelicitous with the adnominal ER in (108b)²¹, although further investigation of both kinds of example is required.

- (108) a. **Is he** himself gonna chair the committee?
 b. # **Is'e** himself gonna chair the committee?

The important observation is that absence of pitch accent on the associate of an adnominal ER does not violate Stress-F and is entirely compatible with the FOCUS SENSITIVE OPERATOR

²⁰ Discussed in von Stechow (1994), Rooth (1996a), Krifka (2004). Example in (108) attributed by von Stechow to Suzanne Tunstall.

²¹ I give an example with a pronoun in the nominative case, since judgments for an adnominal ER in object position (cf. ix) with a pronominal associate are already uncertain (Bickerton 1987, Siemund 2000). Siemund (2000) notes that there is only one attested example in the BNC corpus. I find the cliticized version (x) more marked.

(ix) ? I'm going to dinner with her herself.
 (x) ?? Even YOU're going to dinner with 'er 'erself.

approach. Together with a principal like Stress-F, the approach does not make predictions about prosodic prominence outside the domain of the associate; in Chapter 5, I show that a range of prosodic patterns are possible.

3.3 *Focus operators operate on propositions only*

We've established that the necessary “interface” properties of focus association hold for the adnominal ER to be a sub-propositional focus operator: a syntactic relationship of c-command between the ER and its focus associate and a mapping between a sub-propositional semantic scope and prosodic prominence. In this section, I consider semantic evidence that the adnominal ER has sub-propositional scope: first, that the empirical facts support it and, second, that our theoretical machinery permits it.

In Section 3.3.1, I argue that the adnominal ER orders its associate relative to alternative individuals; previously, it has been assumed that the adnominal ER affects (whether directly or indirectly) an ordering of alternative propositions.

Cunningham (2009) arrives (independently) at similar conclusions about the meaning and scope of adnominal ERs. She rejects focus-based accounts of them because, she argues, the accounts contrast propositions, not individuals. In Section 3.3.2, I review Rooth's (1992) alternative semantics for focus and apply them to the adnominal ER.

Finally, in Section 3.3.3, I consider and reject a syntactic movement approach to the focus of adnominal ERs in the spirit of Wagner (2006). I argue that contrasts in acceptability judgments between *himself* and post-nominal *only/alone* can be accounted for if postnominal *only/alone*—but not *himself*—operates on propositions.

3.3.1 *The adnominal ER ranks individuals, not propositions*

To begin, we consider the pragmatics of adnominal ERs. In many early analyses of the adnominal ER, *himself* is said to convey the remarkability, expectancy or likelihood of a

proposition (e.g. Edmondson & Plank 1978, Primus 1992, Kemmer 1995). If this is correct, the adnominal ER must operate on entire propositions.

Descriptively, Edmondson & Plank (1978) and most subsequent authors have characterized the adnominal ER as having two opposite readings: a “surprise” reading and a “non-surprise” reading. On a surprise reading, the associate is the least likely alternative to be true of the predicate. For example, in (109), the director is the least likely individual to attend the informal meeting.

- (109) The director himself attended our (informal) meeting.
(cf, Siemund 2000)

Let’s informally assume an alternative set *Alt* (110a) ordered by a relation *R* (110b).

- (110) a. Alternative set *Alt*: (to be revised)
 { ...
 The regional managers attended the meeting $>_R$
 The general manager attended the meeting $>_R$
 The associate director attended the meeting $>_R$
 The director attended the meeting $>_R$ }
 b. Ordering relation *R*: Likelihood

The proposition expressed, that the director attended the informal meeting, is most lowly ranked among alternative propositions, according to likelihood. In the next example (111), the proposition expressed, ‘The king wore a crown’ is ranked highest among alternative propositions, according to likelihood. Note that I could have changed the polarity and set the ordering relation as unlikelihood and set the proposition ‘the king wore a crown’ as most lowly ranked. The point is that (111), a “no-surprise” example, involves a scale with opposite ordering from the so-called “surprise” example in (110).

- (111) The king himself wore a crown.
(Eckardt 2001)

- (112) a. Alternative set *Alt*: (to be revised)
 { **The king** wore a crown
 The duke wore a crown $>_R$
 The courtier wore a crown $>_R$
 The court jester wore a crown $>_R$
 The serf wore a crown $>_R$
 ...}
 b. Ordering relation *R*: Likelihood

Eckardt (2001) suggests that (111) and other “no-surprise” examples in the literature are most felicitously realized with a contrastive topic prosody: in English, typically a rise-fall-rise on *himself* and a rise-fall somewhere on the predicate (ToBI annotations in 113). The context for (113) is such that a ‘double focus’ or ‘contrastive focus’ configuration is possible: there are salient discourse antecedents of the form ‘x wore y’.

- (113) (The archbishop was easy to spot, due to his mitre. The Lords wore shining helmets...)

H*-L (H%)	H* L-L%
The king himself	wore a crown

I do not doubt that this prosody and corresponding semantics are possible in utterances with the adnominal ER; In Chapter 5, we will examine several naturally-occurring examples. It is a mistake to assume that all instances of “no-surprise” can be reduced to contrastive topic or double focus, however. The same prosody is available for “pair-list” examples for which the last example is, in fact, a surprise, as in (114).

- (114) (The archbishop was easy to spot, due to his mitre. The Lords wore shining helmets...)

H*-L (H%)	H* L-L%
The king himself	wore a dunce cap

I will suggest that the surprise/non-surprise distinction is not a useful one. More often in examples of the adnominal ER, it is individuals which are ranked, and they are ranked on a scale unrelated to the predicate.

For example, the context of (115) is such that all salient individuals (e.g. all individuals present in central Rome, including the Vatican) are equally likely to perish in the rubble. Assuming that John, who lives in an apartment at the centre of Rome, and the Pope were both near the epicenter of the earthquake, the propositions ‘The Pope perished in the rubble’ and ‘John perished in the rubble’ should be equally ranked by an ordering relation of likelihood.

- (115) *A very powerful earthquake struck the centre of Rome and the Vatican*
The Pope himself perished in the rubble. (Bergeton 2004)

Himself is felicitous in (115) because we agree that the Pope is highly ranked according to an ordering relation of sociological importance. (We can disagree on whether he *should* be sociologically important, but it is hard to deny that he is.) The ordering may be complete, with every salient individual ranked with respect to every other, or partial, with only the Pope singled out and ranked relative to the rest.

- (116) a. Alternative set *Alt*:
 { **The Pope** $>_R$ John, Mary, Brianna, Zhiguo, ... }
 b. Ordering relation *R*: sociological importance

What *himself* contributes in all of these cases is a kind of “big shot” inference, exemplified by the preponderance in the literature of kings, presidents, celebrities and the like, individuals who tend to be highly ranked culturally in all situations. *Himself* is also felicitous with individuals ranked more lowly on an absolute cultural scale, but high nonetheless in certain situations (cf. also Siemund 2000:134).²² The context of (117) is such that all employees are equally likely to testify; we are told that all will appear before the grand jury. The custodians are

²² Golde (1999) makes a similar claim with respect to the “noteworthiness” of the associate with respect to more “humble” but not specifically identified individuals.

highly salient, however, in virtue of the nature of the alleged crime. In addition to the “double focus” prosody, this example is also felicitous with an “all-new” prosody: pitch accents on each lexical word.

- (117) *All of the employees of that company will have to appear before the grand jury, where they will be asked what they know about the alleged illegal trash disposal.*

The custodians themselves will testify late Thursday afternoon [...] (Baker 1995:79)

- a. Alternative set:
 { **the custodians** $>_R$ the accountants, management, ... }
- b. Ordering relation R : Likelihood to have information about the trash disposal

Occurrences of the adnominal ER presuppose that the associate is most highly ranked by some contextually-salient ordering relation: the putative “surprise” and “non-surprise” examples (109) and (111) are therefore better represented with the alternative sets and ordering relations in (118) and (119), respectively.

- (118) a. Alternative set *Alt*:
 { **The director** $>_R$
 The associate director $>_R$
 The general manager $>_R$
 The regional managers $>_R$
 ... }
- b. Ordering relation R : rank in the company hierarchy

- (119) a. Alternative set *Alt*:
 { **The king**
 The duke $>_R$
 The courtier $>_R$
 The court jester $>_R$
 The serf $>_R$
 ... }
- b. Ordering relation R : social status

As Siemund (2000) observes²³, the whatever sense of remarkability or likelihood claimed to exist in such examples persists even when the adnominal ER is removed.

(120) The diREcTor attended the meeting!

This suggests that remarkability or likelihood, when present, is not due to the adnominal ER; and the ordering relation contributed by *himself* need not be related to the predicate. At the same time, it is also the case that nothing prevents the ordering relation from being likelihood of the proposition or, more accurately, the likelihood that the predicate holds of the individual. In fact, in the absence of other salient orderings, likelihood may be the easiest ordering to accommodate.

Suppose I utter (121) out of the blue.

(121) **Mary herself** attended our meeting at work today.

You do not know anything about the person named Mary nor the meeting I am referring to, but are willing to accommodate the existence of said individual and meeting. Not knowing Mary, you do not have any already established way of relating her to other individuals (e.g. being my boss, the head of the company or an A-list celebrity). You might reasonably accommodate an ordering relation of importance at my place of work. You might also reasonably accommodate an ordering relation based on the likelihood of attending the meeting.

- (122) a. Alternative set *Alt*:
 {Mary $>_R x_1 \dots x_2$ }
 b. Ordering relation *R1*: importance at my place of work
 c. Ordering relation *R2*: likelihood of attending our meeting

Alternatively, you as a listener may choose not to accommodate an ordering relation at all. In (123), adapted from Cunningham (2009), we do in fact know the individuals Polly, Emma and

²³ Siemund (2000:129)

Simon. Knowing that Polly is a faculty member makes it easy to accommodate Polly's high position on an ordering of academic rank. On the other hand, there is nothing salient in the context which give us reason to rank Simon above Polly or Emma, making an utterance with *Simon himself* considerably more marked.

- (123) *Emma and Simon are both graduate students, while Polly is faculty member. Emma, Simon and Polly have similar levels of expectations, and it is usually the case that if one of the three approve of some work, the others are equally satisfied.*
- a. The undergraduate semantics research assistant has been doing a lot of really good work lately. Polly herself has commended it.
 - b. The undergraduate semantics research assistant has been doing a lot of really good work lately. #Simon himself has commended it.

3.3.2 The adnominal ER within an alternative semantics for focus

On an alternatives semantics for focus, the semantics of focus are defined recursively. The constituents of (110) have the focus semantic values in (111). The focused constituent *Bill* has a focus semantic value which is the set of individuals, and this is inherited compositionally to the VP and the clause.

- (124) John introduced [Bill]_F to Sue.

- (125) Focus semantic values for constituents in (110) (cf. Rooth 1996b)

$$\begin{aligned}
 \llbracket \text{Bill} \rrbracket^f &= E, \text{ the set of individuals} \\
 \llbracket \text{John} \rrbracket^f &= \{j\}, \text{ the unit set } \llbracket \text{John} \rrbracket^o \\
 \llbracket \text{Sue} \rrbracket^f &= \{s\}, \text{ the unit set } \llbracket \text{Sue} \rrbracket^o \\
 \llbracket \text{introduce} \rrbracket^f &= \{\text{introduce}\}, \text{ the unit set } \llbracket \text{introduce} \rrbracket^o \\
 \llbracket \text{VP introduced Bill}_F \text{ to Sue} \rrbracket^f &= \{\lambda x. \text{introduce}(x, y, s) \mid y \in E\} \\
 &\quad \text{the set of properties of the form 'introducing } y \text{ to Sue'} \\
 \llbracket \text{s John introduced Bill}_F \text{ to Sue} \rrbracket^f &= \{\text{introduce}(j, y, s) \mid y \in E\} \\
 &\quad \text{the set of propositions of the form 'John introducing } y \text{ to Sue'}
 \end{aligned}$$

The focus is interpreted by focus interpretation operator \sim (Rooth 1992), which fixes the scope of focus and determines a discourse antecedent k . For example, the operator \sim in (112b) requires a discourse antecedent k of the form *John introduced x to Sue*.

- (126) a. [John introduced **Tom** to Sue]_k
 b. No, [[_S John introduced [**Bill**]_F to Sue] $\sim k$]

Formally, discourse antecedence is licensed by entailment: a discourse antecedent k ENTAILS the focus semantic value with existential closure f (cf. Rooth 2008, Schwarzschild 1999). (112a) is a possible discourse antecedent for (112b) because it entails the focus semantic value of (112b) with existential closure, namely (112').

- (112') a. $k = \text{introduce}(j, t, s)$ 'John introduced **Tom** to Sue'
 b. $f = \exists x. \text{introduce}(j, x, s)$ 'John introduced **someone** to Sue'

In principle, a focus operator can apply to any constituent. (113) is one version of an often cited example of focus induced interpretation below the level of the clause, due to Rooth (1985).

- (127) a. An American farmer was talking to a Canadian farmer.
 b. An [[_{NP} American_F farmer]₃ \sim 4] was talking to a [_{NP} Canadian_F farmer]₄ \sim 3]

The scope of each focus is fixed by the focus interpretation operator \sim at the level of the NP. In each case \sim requires an antecedent that entails $\exists x. \text{farmer}(x)$, which is satisfied by the other NP. Note that no propositional antecedent is required; indeed there is no discourse antecedent of the form 'A P farmer was talking to a Q farmer'.

The scope of focus for an adnominal ER is fixed at the level of DP.

- (128) a. The Pope himself perished in the rubble.
 b. [[the Pope]_F \sim 2] himself] perished in the rubble.

The focus interpretation operator \sim requires only a discourse antecedent of the form ' x ' where x is an individual. This requirement is notably weak, and we'll consider an objection to its

weakness and the role of focus below. What is important is that (128) is felicitous without a propositional discourse antecedent of the form ‘*x* perished in the rubble’. In other words, (103) is compatible with a situation in which no one else has perished in the rubble and is felicitous in an utterance context in which perishing in the rubble is not salient for any other individual.

This behavior is in contrast to the focus particle *even*, often contrasted with the adnominal ER, which is syntactically a clausal operator and does in fact require a propositional discourse antecedent. (129) requires a discourse antecedent of the form ‘*x* perished in the rubble’.²⁴

²⁴ Mats Rooth (p.c.) notes that in (xii), similar to an example due to Ruth Kempson, that the most salient interpretation does not require that the predicate, *getting back to building treehouses*, be salient and that it is most naturally realized without deaccenting on the predicate.

- (xi) All the children are getting back to their routines after their long convalescence.
(Even) little Johnny has (even) started building treehouses (even).

Jackendoff (1972) attributes a similar example (xii) to Steve Anderson.

- (xii) The results of today’s games will be remarkable: Harvard will (even) defeat Loyola (even).

A couple of analyses suggest themselves to me. First, we might take *even* to require a propositional discourse antecedent in such cases as well, the only difference being that the associate happens to be the entire clause, with paraphrases like (xii-xiv). Johnny’s treehouse building is ranked as the least likely example of the children returning to their routines, and Harvard defeating Loyola the least likely result of the day’s games.

- (xiii) It is even the case that little Johnny has started building treehouses
[even [little Johnny has started building treehouses]_F ~ 3]
(xiv) It is even the case that Harvard will defeat Loyola.
[even [Harvard will defeat Loyola]_F ~ 3]

Alternatively, it might be that there are two focus associates (see Chapter 4, Section 2.6 on double focus), as in (xv-xvi), and we accommodate pair-list alternatives.

- (xv) It is even the case that little Johnny has started building treehouses
[even [little Johnny]_F [has started building treehouses]_F ~ 3]
Salient alternatives: {Aimee is reading Nancy Drew > Betty is playing in her band >
Charlie is going to the gym > Johnny has started building treehouses}
(xvi) It is even the case that Harvard will defeat Loyola.
[even [Harvard]_F [will defeat Loyola]_F ~ 3]
Salient alternatives: {Cornell will win against Syracuse > Florida will tie with Michigan
> NYU will defeat Boston > Harvard will defeat Loyola}

- (129) a. Even the Pope perished in the rubble.
 b. [Even [the Pope_F perished in the rubble] ~3]

3.3.3 The adnominal ER within a covert movement analysis of focus

We saw in Section 2 that the adnominal ER shares many properties with post-nominal focus operators *alone* and *only*, including syntactic scope over a DP associate. I have argued that the adnominal ER also has semantic scope over the DP. Yet postnominal *alone* and *only* are thought to have semantic scope over a clause. In other words, they may share syntactic position, but semantically *himself* operates on individuals and *alone* and *only* operate on propositions. How can we account for this difference.

Kadmon (2001) shows that an in-situ (i.e. non-movement) theory of focus association, an operator like *only* which is a nominal or VP modifier syntactically, may nonetheless be semantically interpreted as a propositional operator.²⁵ This is achieved on the alternatives semantics theory using domain restriction, as in Rooth (1992). Quantifiers like *everyone* or *only* have a pragmatically constrained domain *C* in a given use (cf. 130). The role of focus, according to Rooth (1992), is to further constrain this quantificational domain. I apply Kadmon's notation to pre- and postnominal *only* and postnominal *alone* in (133).

- (130) a. Prenominal *only* quantifying over propositions
 $P (\llbracket \text{only } [\alpha] \rrbracket^g) = \{ w \in W: \text{for all } d \in C \subseteq D_e \text{ if } w \in P(d) \text{ then } d = \llbracket \alpha \rrbracket^g \}$
 where *C* is contextually given set of (sets of) individuals
- b. Postnominal *only* quantifying over propositions
 $P (\llbracket [\alpha] \text{ only} \rrbracket^g) = \{ w \in W: \text{for all } d \in C \subseteq D_e \text{ if } w \in P(d) \text{ then } d = \llbracket \alpha \rrbracket^g \}$
 where *C* is contextually given set of (sets of) individuals
- c. Postnominal *alone* quantifying over propositions
 $P (\llbracket [\alpha] \text{ alone} \rrbracket^g) = \{ w \in W: \text{for all } d \in C \subseteq D_e \text{ if } w \in P(d) \text{ then } d = \llbracket \alpha \rrbracket^g \}$
 where *C* is contextually given set of (sets of) individuals

²⁵ Kadmon's examples illustrate VP-*only*, but she considers (pronominal) NP-*only* to be parallel. She does not discuss postnominal *only* or *alone*.

Certain distribution facts suggest that an in-situ analysis of postnominal *only* and *alone* cannot be correct. In the discussion of the syntax of adnominal ERs in Section 2, we reviewed several naturally-occurring examples of the adnominal ER in genitive constructions (44-52); postnominal *alone* and *only* in the same contexts were ungrammatical. To those examples, we can also add genitive constructions with ellipsis (cf. 131-133).

- (131) *Each node constructs a node ID based on
the selected CH node's information and **the node itself's**.*
An Efficient Aggregation and Routing Algorithm Using Multi-hop Clustering in
Sensor Networks in Computational Science-ICCS 2004, p. 1203 By Marian Bubak,
Geert Dick van Albada
- (132) *StateSec actually managed to steal the Manties' Foreign Office key.*
Not **the Foreign Secretary herself's**, but they did get the departmental key.
from At All Costs by David Weber (2005), Chapter 14
- (133) *Most High was the victim of a villainous plot concocted by some obscure authority
whose ubiquity rivaled **the Lord Himself's**
-- but trying to save the child would have been unthinkable.*
Olivo Oliva p. 60 By Philippe Poloni, David Homel (2006)

In the same environment, postnominal *only* and *alone* are ungrammatical.

- (134) *The node ID's information comes from **the node alone/only 's**.
- (135) *They stole Manties' Foreign Office key.*
* Not **the Foreign Secretary alone/only 's**, though.
- (136) * ... some obscure authority whose ubiquity rivaled the Lord **alone/only 's**.

Discourse antecedence and quantifier domain restriction are both pragmatic mechanisms and meant to account for judgments of infelicity. If the unacceptability of (53-54) and (134-136) were due to infelicity, we should expect them to improve with sufficient contextual support. As far as I can tell, they do not. The examples are not merely infelicitous, but ungrammatical, and I will argue that the ungrammaticality is a result of syntactic movement violation.

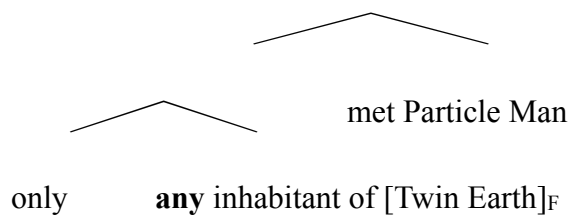
Although contextual domain restriction was introduced in Rooth (1985,1992) precisely as an alternative to focus movement, movement and domain restriction are not incompatible. Wagner (2006) presents a theory of covert focus movement motivated by the distributional pattern of negative polarity items (NPIs) like *any*. The generalization is the following.

- (137) *Only* licenses NPIs in its scope (its second argument), but not in its syntactic restrictor (its first argument)

The generalization assumes that *only* (and *alone*) takes two arguments. The two arguments will be the focus associate (the syntactic restrictor) and a property (the syntactic scope). For example, in the case of DP-*only*, the entire DP is the syntactic restrictor and no NPIs are permitted (cf. 138-139). The predicate is the syntactic scope and NPIs are licensed (cf. 140). (Examples 138-139 also demonstrates that the correct generalization for the distribution of NPIs cannot be that NPIs are licensed simply on the unfocused part of a sentence involving *only*, cf. Horn 1996, Beaver and Clark 2003.)

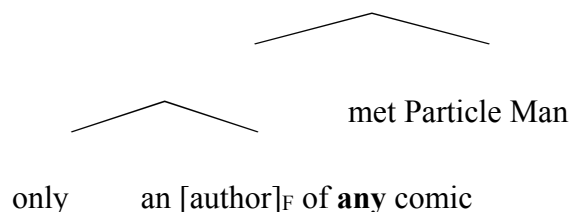
- (138) **Only any inhabitant of Twin Earth met Particle Man.*

LF:



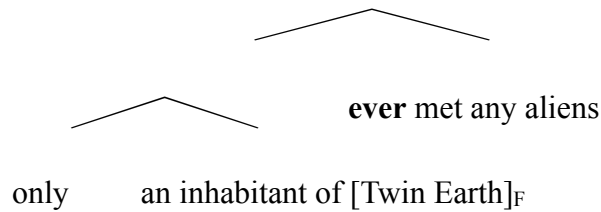
- (139) **Only an author of any comic met Particle Man.*

LF:



(140) Only an inhabitant of Twin Earth ever met any aliens

LF:



In the case of VP-*only*, a constituent must move to the top of the clause in order to provide *only* with its first argument, as in (141).

- (141) a. John only gave any kale to *his* friends.
 [only *his* friends] [λx . John gave **any** kale to x]
 b. John only gave any *kale* to any of his friends.
 [only *kale*] [λx . John gave x to **any** of his friends]

Using the NPI facts as a test for which constituents form part of the syntactic restrictor and which the syntactic scope, Wagner observes several movement constraints, including bans on moving a low object across a high object in double object constructions (cf. 142), and moving a head without its complement (cf. 143-144).

- (142) a. She only gave *her* student any funding.
 [only *her* student] [λx . she gave x **any** funding]
 b. * She only gave any student *summer* funding
 [only *summer*] [λx . she gave **any** student x funding]

- (143) *John only *cut* any vegetables.
 [only *cut*] [λx . John x **any** vegetables]

- (144) *John only drove *through* any park.
 [only *through*] [λx . John drove x **any** park]

Postnominal *only* and *alone* are similarly ungrammatical in genitive constructions because they are subject to a movement constraint: the focus associate is “trapped” within an island, the subject of a genitive. *Himself* remains grammatical in this context because the adnominal ER is

not a propositional operator: its argument is not required to move and therefore violates no constraints on movement.

- (145) a. Claudia's father was **the emperor himself's** dearest friend.
[himself the emperor] [$\lambda x. x$]
b. * Claudia's father was **the emperor only/alone 's** dearest friend.
[only/alone the emperor] [$\lambda x. \text{Claudia's father was } x\text{'s dearest friend}$]
- (146) a. **The song itself's** name is available in-video.
[itself the song] [$\lambda x. x$]
b. * **The song only/alone 's** name is available in-video.
[only/alone the song] [$\lambda x. x\text{'s name is available in-video}$]
- (147) a. **the music itself's having** been tampered with...
[itself the music] [$\lambda x. x$]
b. * **the music only/alone 's having** been tampered with...
[only/alone the music] [$\lambda x. x\text{'s having been tampered with}$]

From a syntactic perspective, the grammaticality of examples like (145a, 146a, 47a) is also consistent with *himself* taking the entire clause as its complement. This would likewise require no movement and therefore violate no constraints on movement (cf. 148).

- (148) [himself] [Claudia's father was the emperor's dearest friend]

Non-quantificational or “scalar” *only* offers a parallel: the distributional pattern of NPIs suggests that scalar *only* takes a clause as its complement. By adding a context that favours a scalar reading of *only*, the previously ungrammatical sentences (cf. 142-144) suddenly become acceptable (cf. 149-151).

- (149) *The department was able to give most students a package with summer and fall funding, but...*
the stingy advisors only gave **any** student *summer* funding
- (150) *The trainees were expected to wash, cut and steam their vegetables, but...*
John only *cut* **any** vegetables.
- (151) To win the race, participants must run up, down and around three hills.
But John **only** ran up or down any hill.

According to the generalization (137), these licensed NPIs cannot be part of a syntactic restrictor, leaving us to conclude that the entire clause is the complement (e.g. [only] [John cut any vegetables]) and that semantically scalar *only* ranks entire propositions. Similarly for *himself*, the entire clause is the complement and semantically it should order propositions.

The problem of course, as we have seen, is that this analysis is at odds with the intuitions described above. It is relatively easy to accommodate an ordering that puts Bill Gates most highly ranked among individuals in (152); it is rather difficult to imagine what ranking might apply to the entire proposition ‘I saw Bill Gates at Starbucks yesterday’ and which propositions it might be ranked among.

(152) I saw Bill Gates himself at Starbucks yesterday.

3.4 A focus sensitive operator has distinct focus sensitive interpretations

3.4.1 Association within the DP

I noted above that the focus semantic value of the associate, namely the set of individuals, will be trivially entailed in any discourse. Although it does not immediately follow from this result that the adnominal ER is not a focus-sensitive operator, it does beg the question whether focus semantics is doing any real work.

A non-trivial example of focus in an adnominal ER construction would involve more than one association with focus within the associate. So far, we’ve considered primarily single-constituent associates and assumed the focus to be co-extensive with its semantic scope (cf. 106b).

Interpretational differences, though subtle, are possible by varying either the scope of focus or the F-marked constituent. For example, (153a,b,c) are string-identical, but vary in meaning. In (153a), *itself* signals that the Burj Dubai is highly ranked while in (153b) *itself* signals that the

construction of Burj Dubai is highly ranked. The two examples differ in both scope of focus and F-marked constituent.²⁶

On an account of the adnominal ER without focus, this contrast could be represented as an attachment ambiguity: *itself* attaches to the DP *the Burj Dubai* in (153a) and to the DP *the construction of the Burj Dubai* in (153b).

- (153) a. A popular blog describes the construction of [[the Burj Dubai]_F ~ k] itself]
 b. A popular blog describes [[[the construction of the Burj Dubai]_F ~ k] itself]
 c. A popular blog describes [[[[the construction]_F of the Burj Dubai] ~ k] itself]

(153c) is a more interesting case because the F-marked constituent occurs is not adjacent to the adnominal. Let's contrast (153c) with (153b). In (153b), the speaker ranks the construction of the Burj Dubai among all other salient entities in the discourse. For example, suppose we are discussing amazing engineering feats that include landing on the moon, digging the Panama Canal and laying the Langed Pipeline. Among these alternatives, constructing the Burj Dubai is most highly ranked according to impressiveness from the speaker's point of view.

In (153c), we predict that the construction of the Burj Dubai is most highly ranked within a smaller alternative set, namely all salient entities of the form 'x of the Burj Dubai'. For example, suppose we are discussing the history of the Burj Dubai. Among the alternatives are the design of the Burj Dubai, the funding for the Burj Dubai, the interior design of the Burj Dubai, and so forth. According to an ordering relation of impressiveness in the view of the speaker, it is the construction of the Burj Dubai which is ranked most highly. 'Landing on the moon', 'digging the

²⁶ Note that the ambiguity is unavailable when the two nominals have different phi-features, as illustrated in (xiv).

- (xvii) a. A popular blog describes the coronation of [[Queen Elizabeth]_F ~k] herself.]
 b. *A popular blog describes [[[the coronation of Queen Elizabeth]_F ~k] herself]
 c. *A popular blog describes the coronation of [[[Queen Elizabeth] _F ~k] itself]
 d. A popular blog describes [[[the coronation of Queen Elizabeth]_F ~k] itself]

Panama Canal’ and ‘laying the Langed Pipeline’ are not possible alternatives because they do not have the form ‘x of the Burj Dubai’.

(156) b. A popular blog describes [[[the construction of the Burj Dubai]_F ~ k] itself]

(154) a. Alternative set

{the construction of the Burj Dubai >_R
 landing on the moon >_R
 digging the Panama Canal >_R
 laying the Langed Pipeline >_R
 ...}

b. Ordering relation *R*: impressiveness according to the speaker

(157) c. A popular blog describes [[[[the construction]_F of the Burj Dubai] ~ k] itself]

(155) a. Alternative set

{the design of the Burj Dubai >_R
 the funding of the Burj Dubai >_R
 the interior design of the Burj Dubai >_R
 ...}

b. Ordering relation *R*: impressiveness according to the speaker

The effect of association with focus in an adnominal ER is present, although admittedly subtle; not all focus-sensitive constructions are truth conditional.

3.4.2 *Scalarity without focus*

Cunningham (2009) proposes a semantics for adnominal ERs without focus semantics, a version of which is given in (156).

(156) $\llbracket \text{himself} \rrbracket = \lambda x_e. \exists R [\exists s [x \text{ is a scalar endpoint on } s^C \text{ with ordering relation } R^C]]. [x=x]$

In particular, Cunningham suggests that because of their scalar behavior, adnominal ERs pattern more parsimoniously with gradable adjectives (e.g. *tall*, *expensive*). According to Cunningham, both the adnominal ER and gradable adjectives map their argument onto a scale: gradable adjectives to a scale of degrees (Kennedy & McNally 2005) and the adnominal ER to a scale of individuals.

As discussed above, many of the associates found in the literature on adnominal ERs (e.g. President, director, king) do belong to easily accessible real-world hierarchies, such as constitutional, corporate or feudal rank. Bayer (1996) notes, however, that there is nothing inherent in proper names that trigger any kind of scale. *Obama* or *British Petroleum*, for example are not inherently scalar. While this suggests that the scalar interpretation in (157-158) must indeed be triggered by the adnominal ER, rather than its associate, the adnominal ER does not contribute to the meaning of a scale of individuals in the same way that *old* or *liberal* contribute a scale of degrees. The adjective *old* invokes a conventionalized scale with degrees relative to chronology (e.g. days, weeks, years); the adjective *liberal* will map individuals to degrees of “liberalness”. While both these scales will be context-sensitive, what counts as a degree for such a predicate is nonetheless highly constrained. The possible rankings for *Obama* and *British Petroleum*, by contrast, are in principle infinite.

- (157) **Obama himself** will sign the treaty.
- (158) **British Petroleum itself** has admitted wrongdoing.
- (159) The **old man** sat in a chair.
- (160) **Meagan** is very **liberal**.

Cunningham’s positive argument for a focus-less semantics of adnominal ERs appeals to parsimony—adnominal ERs pattern with gradable adjectives in some respects—and my initial reply has taken roughly the same tact: adnominal ERs pattern do *not* pattern with gradable adjectives in some other respects. My second reply is also an appeal to parsimony. Several scalar particles in English and other languages are at least homophonous with focus particles, and have been analyzed as such. The association with focus behavior of the scalar additive *even* has long been noted (e.g. Jackendoff 1972), and *only*, another classic focus particle, has been studied on its scalar interpretation, as well (e.g. Bayer 1996; van Rooij 2002; Klinedinst 2004; Reister 2006).

- (161) Even [the vegetarian]_F tried a smoked meat sandwich scalar *even*
- (162) *After ten years at university* (Klinedinst 2004) scalar *only*
 Bill only has a [master's degree]_F

From these cases, it is equally justified to lump the adnominal ER with the set of focus particles as with gradable predicates. More likely, the semantics of focus and the semantics of scalar constructions are not mutually exclusive, and one can imagine an ambitious project that seeks to unify the semantics of scalar focus particles like *even* and *only* with gradable adjectives. The important point is that a scalar nature of the adnominal ER does not constitute strong evidence in favour of or against a focus operator approach to the adnominal ER.

3.5 *Focus operators may interact*

The next property concerns the behavior of the adnominal ER with other focus operators. König (1991), anticipating worries of Cunningham (2009), observes that the adnominal ER may co-occur with other focus operators, as in (163-164), and that both operators may share the same associate.

- (163) **Even** the victim of the accident **himself** is content with the settlement.
 (164) **Only** the President **himself** can make this decision.

Hole (2002) is concerned about the configuration in which the adnominal ER itself is the focus associate of another operator. Noting the infelicity of (165), he objects that examples like (163) and (164) with focus on *himself* ought to be, but are not, similarly infelicitous.

- (165) #The artist will **only** come **TOO**.

In all of these respects, however, the adnominal ER patterns like familiar focus sensitive operators. Krifka (1992) recognizes five different multiple focus configurations illustrated in (166). Note that in all but (166a) two focus sensitive operators co-occur; in (166d), two

operators share the same focus associate; in (166e), one operator is itself the associate of a second operator; and all examples in (166) are perfectly grammatical and interpretable.

- (166) a. John only₁ introduced [Bill]_{F1} to [Sue]_{F1}
 b. Even₁ [John]_{F1} drank only₂ [water]_{F2}
 c. John even₁ [only₂ drank [water]_{F2}]_{F1}
 d. John even₁ only₁ drank [[water]_{F2}]_{F1}
 e. John even₁ drank [only₂]_{F1} [water]_{F2}

The infelicity of (165) may be attributed to the structural principle discussed above that certain focus sensitive operators including *only* must associate with a lexical constituent in their c-command domain. *Only* attaches to the VP, while *too* attaches to the clause. *Too* is therefore not an eligible associate for *only* because it is not within the c-command domain of *only*. Pragmatically, it is also unclear what (165) *could* mean. Intuitively, *only* and *too* are contradictory: *only* rules out other alternatives from a set while *too* presupposes that there is at least one other alternative.

4 The extended focus sensitive operator approach

4.1 Reconciling approaches to the adnominal ER

Let's now return to the two formal semantic approaches to the meaning of the adnominal ER: the FOCUS SENSITIVE OPERATOR approach and the FOCUSED ASSERTION OF IDENTITY approach. The representations of the reports are repeated in (167) and (168), respectively.

- (167) FOCUS SENSITIVE OPERATOR
 [[[the Provost]_F ~] himself] will chair the committee

- (168) FOCUSED ASSERTION OF IDENTITY
 [the Provost [himself]_F will chair the committee] ~]

Both accounts recognize that the adnominal ER does not contribute anything to the truth-conditions of an utterance. Semantically, the adnominal ER is a syntagorematic operator, asserting only the identity function (cf. 169-170).

$$(169) \llbracket \text{himself} \rrbracket = \lambda x_e.[x=x]$$

$$(170) \llbracket \text{himself} \rrbracket(\text{the Provost}) = \lambda x_e.[x=x](\text{the Provost}) = \{\text{the Provost}\}$$

As we affirmed in Section 3.5, a focus operator can itself be focused by another focus operator. This is the premise of the combined approach: the adnominal ER is a focus operator which itself may be focused. The representation in (167) is nothing more than a special case of (160), with one focus operator is nested within another. In (171), *the Provost* is both the scope and the focus associate of the focus sensitive operator *himself*; *himself*, in turn, is the focus associate of a focus operator having scope over the entire utterance.

$$(171) \text{ FOCUSED FOCUS SENSITIVE OPERATOR}$$

$$[[[\text{the Provost}]_{F1} \sim 1] \llbracket \text{himself} \rrbracket_{F2} \text{ will chair the committee}] \sim 2$$

In section 4, I discuss the predictions of this account. I claim that the pragmatic contribution of the adnominal ER is the ordering relation *R*. The overall prosodic pattern of a sentence containing an adnominal ER is not pre-determined by the adnominal ER, but instead varies with context according to available discourse antecedents.

4.2 Alternatives and discourse antecedence

Semantically, the adnominal ER is a function of type $\langle ee \rangle$, from individuals to individuals; it maps an individual to itself. Alternatives to this function will be other functions of type $\langle ee \rangle$, i.e. relational predicates such as *assistant to*, *nominee of* and *wife of* (172a). For instance, if the argument of *himself* is *the Provost*, the focus semantic value of *the Provost himself* will be the set of functions *u* of type $\langle ee \rangle$ of the form '*u*(Provost)' (172b).

$$(172) \text{ a. } \llbracket \text{himself}(x) \rrbracket^f = \llbracket \lambda x_e.[x=x] \rrbracket^f = \{\lambda x.(x) \mid u \in D^{\langle ee \rangle}\}$$

the set of functions from *x* to other individuals

$$\text{b. } \llbracket \text{himself}(\text{the Provost}) \rrbracket^{f,g} = \{\text{ASSISTANT-TO}(\text{the Provost}), \\ \text{NOMINEE-OF}(\text{the Provost}), \\ \text{WIFE-OF}(\text{the Provost}) \\ \text{etc.}\}$$

The focus operator which associates with the adnominal ER attaches at the clause level and so requires a clausal discourse antecedent. In (173b), *himself* requires a salient discourse antecedent that that someone related to the Provost will chair the committee. This discourse requirement is satisfied by (173a), which entails the focus semantic value of (173b) with existential closure (cf. 174).

- (173) a. [The Provost's assistant will chair the committee]₂
 b. No, [[the Provost [himself]_F will chair the committee] ~₂
- (174) a. discourse antecedent *k*
 'The Provost's assistant will chair the committee'
 b. focus semantic value of (173b) with existential closure *f*
 $\exists u_{ee}.\text{chair}(\text{committee}, u(\text{Provost}))$
 'Someone standing in a relation to the Provost will chair the committee'

The prosody of (173) is governed by Stress-F. The adnominal ER must be most prominent within the phonological domain corresponding to the clause. In Chapter 5, we'll examine several naturally occurring examples of this configuration from the web harvested corpus speech. This is where the story ends for the FOCUSED ASSERTION OF IDENTITY approach: the adnominal ER is focused by a clause-level operator and the adnominal ER is realized with greatest prominence in the corresponding phonological domain.

For the FOCUS SENSITIVE OPERATOR approach, this is just the beginning of the possibilities. In addition to narrow focus on the adnominal ER, the adnominal ER may also occur with focus on an entirely different constituent, or it may occur within 'broad-level' or 'all-new' focus on the entire clause. The clause-level focus operator in (175), for example, is trivially licensed by any antecedent of sentential type. In (176), it does not need to be salient in the discourse that someone, either related to the Provost or not, will chair a committee.

- (175) a. What's new?
 b. [The Provost himself is going to chair some committee]_{F1} ~₁

- (176) a. discourse antecedent k
 any utterance having semantic type <st>
 b. focus semantic value with existential closure f
 $\exists p.p$
 any utterance having the semantic type <st>

This is possible on the FOCUS SENSITIVE OPERATOR approach because focus above the DP is independent of the focus *within* the adnominal ER. We represent this as two nested focus operators ~ 1 and ~ 2 .

- (176') a. What's new?
 b. [[[The Provost]_{F2} ~ 2] himself] is going to chair some committee]_{F1} ~ 1

The domain corresponding to the clause will be realized with “default” prosody, in the sense that phrasing and prominence will be determined by regular phonological principles. Most words, including the adnominal ER, will be realized with pitch accents; the adnominal will be neither most prominent, nor deaccented.

I provide a full description of several focus configurations and their prosodic patterns, with naturally-occurring and constructed examples, in Chapter 5. The important point for now is that the FOCUSED ASSERTION OF IDENTITY approach predicts a single, clause-level focus configuration and corresponding prosodic pattern.²⁷ In fact, focus on the adnominal ER is just one of several clause-level focus configurations, although it is a configuration that occurs frequently enough that the FOCUSED ASSERTION OF IDENTITY approach takes it to be the only or the paradigmatic case. It is this configuration with focus on the adnominal ER itself that led some researchers to conclude that the adnominal ER is “invariably in focus and therefore typically stressed” (König & Gast 2006:223), and that alternatives evoked “must have something to do with the asserted value” (Gast 2006:39).

²⁷ The FOCUSED ASSERTION OF IDENTITY approach predicts two focus configurations if we allow double focus or contrastive topic configurations with one focus on the adnominal ER and one focus elsewhere in the clause. Both foci must associate with a clause level operator.

Descriptively, the relationship between an associate X and an alternative Y have been characterized in the following ways (from König & Siemund 2000a,b,c).

- (177) a. X is more significant than Y in a specific situation;
b. X has a higher position than Y in a hierarchy;
c. X is the center of perspective/narration/subject of consciousness;
d. Y is defined in terms of X.

On the FOCUSED ASSERTION OF IDENTITY approach, focus on the adnominal ER is used to account for all such cases, while on the extended FOCUS SENSITIVE OPERATOR approach, focus on the adnominal ER accounts for only a subset. In the rest of this chapter, I'll compare how the two approaches carve up this descriptive pie.

Recall that on the extended FOCUS SENSITIVE OPERATOR approach, the adnominal ER contributes a context-sensitive ordering relation *R*. It is this relation which is responsible for König & Siemund's first two descriptions: the ordering relation depends on the context (cf. 177a) and the particular ordering may or may not be sociologically salient (cf. 177b).

When the ordering relation is based on a real-world hierarchy (cf. provosts, popes and presidents), we get a "big shot" reading. This reading is prevalent in the literature, I believe, because the importance of a provost, pope or president is easy to accommodate. Other examples, I suggested, will require accommodation on the part of the hearer if s/he does not know the individual referred to be the associate or does not have access to salient ordering relation. In (121), repeated from (124), we may need to accommodate that there is an individual Mary and that she is highly ranked on some scale salient to the speaker. Alternatively, we may choose to challenge the use of the adnominal ER. In (178b), repeated from (125), we may judge the utterance infelicitous or explicitly request more information in order to accommodate it: e.g. "What do you think is so significant about Simon?".

- (178) Mary herself attended our meeting at work today.

- (179) *Emma and Simon are both graduate students, while Polly is faculty member. Emma, Simon and Polly have similar levels of expectations, and it is usually the case that if one of the three approve of some work, the others are equally satisfied.*
- a. The undergraduate semantics research assistant has been doing a lot of really good work lately. Polly herself has commended it.
 - b. The undergraduate semantics research assistant has been doing a lot of really good work lately. #Simon himself has commended it.

Creswell (2002) argues the similar point (cf. also Leskosky):

“For the use of an intensive [= adnominal ER] to be felicitous, the addressee must be able to infer the speaker’s beliefs about the referent’s prominent status within the discourse model. In the participants’ shared understanding of the discourse context, certain entities are more prominent than others. The speaker assumes that the addressee shares (or can understand) the belief that the entity referred to with an intensive is prominent and understands the criteria [= ordering relation] on which this prominence depends.” (Creswell 2002:33)

She offers two examples, taken from a written corpus, with inanimate NPs. In the absence of some salient or easily accommodated ordering relation that ranks zinnias above the other flowers mentioned, (180) is infelicitous. In (181), the rocky beach and ocean are both visually salient, she observes, but only the sea is independently “conceptually prominent”. Alternatively, we might say that the catalogue writer simply wanted to emphasize the sea by singling it out as most prominent in the scene.

- (180) Both annuals and perennials are commonly planted in cut flower gardens. Some recommended annuals include aster, cornflower, cosmos, dianthus, gomphrena, marigold, scabiosa and **zinnia (#itself)**
- (181) *Catalog advertisement illustrated with pictures of a model wearing woolen sweaters, posed on a rocky beach with the ocean in the background*
- a. Enduring elements / Hardy texture, neutral tone, muted pattern, a sense of the rocky beach and of **the sea itself**
 - b. ?? Enduring elements/ Hardy texture, neutral tone, muted pattern, a sense of the sea and of **the rocky beach itself**

Although real-world hierarchies are easily accommodated, it is ultimately the speaker who determines the ordering relation in a given utterance. Returning again to provosts, popes and presidents, suppose that I have been invited to a large black-tie affair which may have many high-profile attendees, including rockstar Lady Gaga and the President of the United States. While you may accommodate my ranking the President most highly in (182a), you will accommodate different rankings in (182b,c). You may care little for politics and feel indifferent about the president, but you can understand in (182a) that I am ranking the President more highly than Lady Gaga and the Pope; conversely, you may have barely heard of Lady Gaga or even dislike her, and have disdain or indifference towards the Pope, but you know that I am signaling that I think Lady Gaga is highly ranked (182b) or the Pope is highly ranked (182c).

- (182) a. I talked to Lady Gaga and the Pope and to the **President himself**
 b. I talked to the President and the Pope, and to **Lady Gaga herself**
 c. I talked to Lady Gaga and the President, and to **the Pope himself**

On the extended FOCUS SENSITIVE OPERATOR approach, the President in (182a) is ranked according to alternatives Lady Gaga and the Pope. The expressions *Lady Gaga* and *the Pope* are also sufficient antecedents for the focus associate *the President*, since they trivially entail that there is some individual (cf. 183). The entire sentence can be realized with “all new” prosody.

- (183) FOCUS SENSITIVE OPERATOR
 a. [[[the President]_{F1} ~ 1] himself]
 b. Antecedent(s): *Lady Gaga, the Pope*
 c. Alternatives: *Lady Gaga, the Pope*

On the FOCUSED ASSERTION OF IDENTITY approach, *himself* must be narrowly focused and the sentence realized with greatest prominence on *himself*. Focus on *himself* evokes alternative relations to the President and requires discourse antecedents of the form ‘I talked to the President’s *u*’, ‘I talked to the *u* of the President’ where *u* is a relational predicate of type <ee> (cf. 184). In (182a), there are no explicit antecedents with this form.

- (184) FOCUSED ASSERTION OF IDENTITY
- a. [I talked to the President [himself]_{F2}] ~2
 - b. Some possible alternatives: WIFE-OF(the President), ASSISTANT-TO(the President)
 - c. Some possible antecedents: ‘I talked to the President’s wife’, ‘I talked to the President’s aide’, etc.

A charitable listener could, however, accommodate a relational predicate such that Lady Gaga and the Pope stand in some relation to the President. Consider the related example in (185). Given our world knowledge that the First Lady stands in the WIFE-OF relation to the President, this relational predicate is easy to accommodate. One might also accommodate the WIFE-OF relational predicate from *Mary*; of course many other relational predicates might also be possible, such as SUBORDINATE-TO, FRIEND-OF.

- (185) I talked to the First Lady/Mary and to the President himself.
- (186) First Lady = WIFE-OF(the President)?
- (187) Mary = WIFE-OF(the President)?, SUBORDINATE-TO(the President)?, FRIEND-OF (the President)?

Returning to our example (182), it is much more difficult, although not impossible, to accommodate a function from *the President* to *Lady Gaga* or *the Pope*. Neither Lady Gaga nor the Pope is a wife of, subordinate to or friend of the President, etc. On the FOCUSED ASSERTION OF IDENTITY approach, it is always necessary to accommodate such a relational predicate (or else challenge the utterance). On the extended FOCUS SENSITIVE OPERATOR approach, we must accommodate this interpretation if and only if the adnominal ER itself is focused, as in (188). This would be realized with greatest prominence on *himself*.

- (188) FOCUS SENSITIVE OPERATOR
- a. [[[the President]_{F1} ~ 1] himself]
 - b. Antecedent(s) for F1: *Lady Gaga, the Pope*
 - c. Alternatives to F1: *Lady Gaga, the Pope*

FOCUS SENSITIVE OPERATOR under focus

- d. $[[[I \text{ talked to } [[\text{the President}]_{F1} \sim 1] [\text{himself}]_{F2}] \sim 2]$
- e. Some possible alternatives to F2: WIFE-OF(the President), ASSISTANT-TO(the President)
- f. Some possible antecedents for F2: ‘I talked to the President’s wife’, ‘I talked to the President’s aide’, etc.

It is these cases of the focusing the focus sensitive operator which correspond to König & Siemund’s fourth relationship, in which alternative entities are defined in terms of the associate by a relationship predicate.

Note that the ordering relation R , which ranks the individuals (in this case an ordering of importance according to the speaker) is distinct from the relational predicate (e.g. WIFE-OF(), SUBORDINATE-TO(), etc.) On the account being defending here, the ordering relation R ranking alternatives must be available in all felicitous utterances of the adnominal ER, while a relational predicate u is required just in case the adnominal ER is focused. In (185), the President is certainly contrasted with his wife; at the same time, however, he is ranked relative to other alternatives according, in this case, to his perceived importance.

Summarizing so far, then, we’ve argued that König & Siemund’s first two relationships (177a,b)—significance in a particular situation and importance in a hierarchy—are accounted for by the ordering relation R , which is contributed by the adnominal ER. The fourth relationship (177d)—defining alternatives in terms of the associate—corresponds to a relational predicate u which is evoked only when the adnominal ER is focused.

The third relationship describes cases in which the set of alternatives includes discourse participants: the speaker (cf.189), the hearer (cf.190) or the subject of consciousness in (191-192). The use of the adnominal ER is contextually sensitive in these cases as well. In (189) the speaker ranks herself most important in the situation (e.g. $\{ \text{speaker} >_R \text{ hearer, others} \}$); in

(190) she ranks the hearer as most important (e.g. { hearer >_R speaker, others}).²⁸ (191-192) are cases of free indirect discourse in which the perspective holder or subject of consciousness is referred to in the 3rd person, and this subject of consciousness ranks herself as most important among alternatives in the situation.²⁹

- (189) John may go shopping while you walk the dog. **I myself** will stay with the kids. (Siemund 2000)
- (190) John may go shopping while I walk the dog. **You yourself** will stay with the kids. (Siemund 2000)
- (191) After dinner Paul proposed to Mary. **She herself** had expected something completely different. (Siemund 2000)
- (192) Jemima guessed that Pompey had chivalrous doubts about leaving her in the gaunt building, with only Tiger, now in a highly restless mood, as company. **She herself** had no such fears. (König & Siemund 2000b)

As in the previous cases, the FOCUSED ASSERTION OF IDENTITY approach assumes that perspectival examples (189-192) require a salient relational predicate *u*. On the extended FOCUS SENSITIVE OPERATOR approach, a salient relational predicate is required only when the adnominal ER is focused; in examples such as (193-195), the adnominal ER is not focus. With the first and second pronoun, or in free indirect discourse, the adnominal ER conveys a kind of deference by partially ranking the associate relative to other individuals.

- (193) Last night, a famous director was in the audience. Guess who will star in his latest production? **I myself**, of course!
- (194) Last night, a famous director was in the audience. Guess who will star in his latest production? **You yourself**, of course!

²⁸ All examples of adnominal ER in the first person are, on this analysis, slightly egocentric and all examples of adnominal ER in the second person are somewhat deferential. I believe that this prediction is correct, but leave a fuller examination to future study.

²⁹ See Gast (2004), König & Gast (2006) and Gast (2006) for discussion of emphatic reflexives in logophoric contexts. These authors recognize that such examples are not straightforwardly captured on the FOCUSED ASSERTION OF IDENTITY approach. In general, they propose that “all DPs occurring in a logophoric environment are interpreted relative to the relevant epistemic validator” (Gast 2004:80), however they do not offer an explicit semantics or paraphrase of examples like (191-192), so I am not certain how their analysis would proceed.

- (195) The night before, a famous director was in the audience. And who did Jane think would star in his latest production? **She herself**, of course!

5 Conclusion

In this chapter, I have presented empirical evidence for the existence of a sub-propositional focus sensitive operator. Several of the properties ascribed to focus sensitive operators in the literature, which could have been arguments against a sub-propositional focus sensitive operator, turn out to be false or require further qualification:

- Some *but not all* focus sensitive operators require c-command over their focus associate (Beaver & Clark 2008)
- The associate of a focus sensitive operator is realized with prosodic prominence *within the phonological domain corresponding to the scope of focus* (cf. Chomsky 1971, Jackendoff 1972, Truckenbrodt 1995, Zubizarreta 1998, Büring 2008, Rooth 2009)
- Rooth's (1992) alternatives semantics for focus allows for focus sensitive operators below the proposition

Additionally, I showed that the adnominal ER *does* appear to c-command its associate, that it *does* show subtle but demonstrable focus-sensitive effects and that it *does* interact with other focus operators.

The adnominal ER is a focus-sensitive operator with DP scope; pragmatically, it contributes a context-sensitive ordering relation *R*, which ranks the associate against the focus alternatives. The interface principle STRESS-F simply requires that the associate is most prominent with its scope, the DP. The meaning contribution of the adnominal ER is therefore independent of sentential focus, contra the FOCUSED ASSERTION OF IDENTITY approach which asserts that the adnominal ER was meaningless unless focused.

On the FOCUSED ASSERTION OF IDENTITY approach, the adnominal ER is said to be invariantly focused; on the extended FOCUS-SENSITIVE OPERATOR approach, focus on the adnominal ER is just one of several focus configurations at the clause level. If correct, the extended FOCUS-SENSITIVE OPERATOR approach predicts that in a sample of naturally-occurring

tokens of the adnominal ER, we should find these different focus configurations and their corresponding prosodic patterns. In the final chapter, I discover these configurations with naturally-occurring speech data from the web and identify them according to the availability of discourse antecedents and the pattern of *f0* contour peaks.

CHAPTER 5
DEBUNKING FOCAL DETERMINISM:
THE CASE OF ADNOMINAL EMPHATIC REFLEXIVES

1 Introduction

Rooth (1992) characterizes a “strong” theory of focus interpretation as one which does not require construction-specific stipulation of focus effects. A “weak” theory stipulates focus effects for each specific construction; an “intermediate” theory stipulates focus effects for some but not all constructions. For example, a weak theory might stipulate a different syntactico-semantic rule or lexical entry that refers to F-marking, focus semantic values or other grammatical reflex of focus, for each focus sensitive construction. Such a theory is considered weak in the sense that a stipulation is descriptive more than explanatory; a weak theory also predicts focus effects to be obligatory. On a strong theory, there is no construction-specific reference to focus, and focus effects are predicted to be optional.

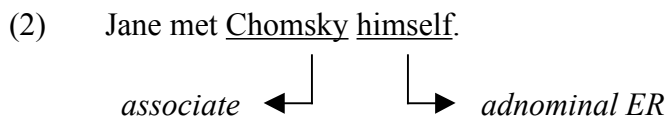
In Chapter 4, I adopted an interface principle STRESS-F (Rooth 2006) which maps semantic scope to a prosodic domain and maps focus associates to prosodic prominence (see also Chomsky 1971, Jackendoff 1972, Truckenbrodt 1995, Zubizarreta 1998, Büring 2008 for similar principles). Assuming such a principle, it is possible to test for semantic focus by examining prosody. I applied some heavy artillery from experimental phonetics and machine learning to test the optionality of focus in cases of “second occurrence focus”. The results of these experiments demonstrated that the apparent optionality is not supported by the prosodic facts, in favor of a weak or intermediate theory of focus.

(1) Stress-F (Rooth 2009)

Let β be an F-marked phrase with scope ϕ . Then the most prominent syllable in the phonological realization of ϕ falls within the realization of β .

In this chapter, I am interested in a slightly different manifestation of the weak or intermediate theory, which I call focal determinism. My question does not concern whether some constructions require there to be a focus (i.e. the optionality of association with focus); rather, my question is whether some constructions are required themselves to *be focused*.

I concentrate on the adnominal emphatic reflexive (ER) introduced in Chapter 4 (cf. 2). Like the debate surrounding second occurrence focus, the semantic arguments for the adnominal ER hang largely on prosodic evidence. One approach to the semantics of the adnominal ER, which I call the FOCUSED ASSERTION OF IDENTITY approach, claims that the adnominal ER must be deterministically focused *because* the adnominal ER is always realized with prosodic prominence.



Analysts within the FOCUSED ASSERTION OF IDENTITY approach differ on how exactly focal determinism arises. One option is that focus is stipulated as a lexical property of the adnominal ER. This is inherent or lexical focus. Another, very clever option is to avoid stipulation by bleaching the semantics and pragmatics of the adnominal ER altogether, such that it lacks any semantic import *unless* it is focused (e.g. Eckhardt 2002, Gast 2006, Ahn 2009). This is obligatory focus.

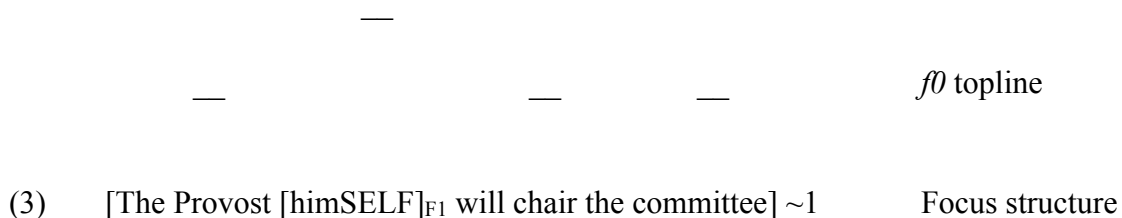
The first option, inherent focus, clearly belongs in a weak theory of focus. One simply stipulates that the adnominal ER is always focused. The second option, obligatory focus, appears to be strong, in the sense that it does not require the semantics of the adnominal ER to refer directly to focus. On the other hand, the second option remains obliquely stipulative in the sense that it no optionality is permitted. Compare other semantically vacuous lexical entries which contribute syntactic or pragmatic information. And if the adnominal ER is just one of

several semantically vacuous, obligatorily focused constructions, it is not clear how we as analysts or learners can distinguish one from the other, beyond semantic type.

My goal in this chapter is to demonstrate empirically that the adnominal ER is not deterministically focused. In order to do so, I must show that other focus configurations may obtain: i.e. that focus is truly optional. And each focus configuration must have distinct pragmatic requirements, and a distinct prosodic realization.

The machine learning approach from Chapter 3 is not well suited for this task. The particular categories of focus configurations are unknown and unlikely to be binary; and because the adnominal ER construction does not require an explicit antecedent, there can be no simple metric for semantic classification. Rather, I propose a careful qualitative investigation.

For this purpose, I represent prosodic prominence with lines, aligned to stressed syllables, to indicate relative height of *f0* peaks, a kind of “*f0* topline” (Eady & Cooper 1986). This representation is of course an idealization and intentionally makes the naïve assumption that pitch accents are exclusively high (H* or L+H*). I do not claim that all pitch accents in such examples ought to be high (although most if not all in the web corpus are in fact high). And, as we confirmed in Chapter 3, focus can of course be realized with non-intonational prosodic cues. The *f0* topline is, however, a simple and relatively objective qualitative method of comparing the prosody of individual tokens of speech.



As for the semantic representation, I apply principles of an anaphoric theory of focus (Rooth 2008b). Focus is interpreted by a focus interpretation operator \sim (Rooth 1992), which both fixes the scope of focus and determines a discourse antecedent. A syntactic constituent is

licensed if and only if its existentially quantified focus semantic value f is entailed by a salient discourse antecedent k . Informally, we can say that the existentially quantified focus semantic value is determined by replacing the F-marked (i.e. focused) constituent with a variable, and existentially quantifying over it. A discourse antecedent may be or may not be explicit.

Let's take a toy example to start. The focus semantic value f of

- (4) [John ate [a cracker]_{F1}] ~ 1

is

- (5) $f = \exists x_{\langle e \rangle}. [\text{John ate } x].$

(5) reads 'there is some thing x (an individual of type e) such that John ate it'. Possible discourse antecedents k for (4) include:

- (6) a. $k = \text{John ate broccoli}$
b. $k = \text{John ate lunch}.$

Both examples entail f . If John ate broccoli, he ate something; and if John ate lunch, he ate something. The antecedent 'John ate lunch' could be made salient indirectly by related utterances, such as (7a,b).

- (7) a. John didn't eat lunch.
b. I heard John ate lunch.

Returning to our example of the adnominal ER, the existentially quantified focus semantic value f of the utterance in (3) is (8). It reads 'there is some relation u such that the u of the Provost will chair the committee' or 'there is some relation u such that the Provost's u will chair the committee'.

- (8) Existentially quantified focus semantic value of (3)
 $f = \exists u. [u(\text{the Provost}) \text{ will chair the committee}]$

Possible discourse antecedents might include the following.

- (9) Possible discourse antecedents for (3)
- a. k = The Provost's assistant will chair the committee
 - b. k = The advisor to the Provost will chair the committee
 - c. k = A representative from the Provost's office will chair the committee

Typically, the scope of focus is the sentence, as in (3); if one's theory permits sub-propositional focus operators, the scope may be smaller, for example at the level of the DP (cf. Section 5). Even in the latter case, the scope must contain at least the adnominal ER and the associate in order for the adnominal ER to contribute to the meaning of the utterance. Minimally, then, the FOCUSED ASSERTION OF IDENTITY predicts that the adnominal ER will be realized with greater prominence than the associate.

For proponents of a FOCUS SENSITIVE OPERATOR analysis (e.g. König 1991, Siemund 2000), the purported prosodic prominence on *himself* is an idiosyncratic mismatch between phonology and syntax, since the focus operator rather than the focus associate is realized with prosodic prominence.

In Chapter 4, I proposed the following solution. The adnominal ER is indeed a focus sensitive operator, with scope at the level of the DP. Its associate is always most prominent within the domain corresponding to this (co-extensive) scope. But the adnominal ER may also be focused, not obligatorily as on the FOCUSED ASSERTION OF IDENTITY approach, but optionally, as one of several possible associates of a propositional focus-sensitive operator.

On this "extended" FOCUS SENSITIVE OPERATOR approach, the adnominal ER is a nested focus sensitive operator: the scope of the adnominal ER is sub-propositional and contained within the scope of a higher focus operator. According to Büring (2008) and Rooth (1996a, 2009), it is this scope relationship which results in second occurrence focus (SOF) (cf. 10).

Recall that in cases of SOF a focus sensitive operator and its associate are repeated (cf. 11) or accommodated as already salient (cf. 12). The principle STRESS-F ensures that the focus

associate of the higher operator (*Petr* in 11 and *the chairman* in 12) will be realized with greatest prominence in the domain corresponding to its larger scope; and the focus associate of the lower operator (*graduate students* in 10 and *younger candidates* in 12) is realized with less prominence relative to the first focus associate, but greatest prominence within its own, smaller domain.

- (10) Configurational second occurrence focus (SOF) (adapted from Rooth 2009)
 [. . . F . . . [[. . . SOF . . .] ~ Op] . . .] ~ Op
- (11) A: Eva only gave xerox copies to the GRADUATE STUDENTS.
 B: No, PETR only gave xerox copies to the graduate students (Partee 1991)
- (11') A: Eva [only [gave xerox copies to the [graduate students]_{F1}] ~1]
 B: [[Peter]_{F2} [only [gave xerox copies to the [graduate students]_{F1}] ~1] ~2]
- (12) A: The provost and the dean aren't taking any candidates other than Susan and Harold seriously.
 B: Even the CHAIRMAN is only considering younger candidates (Rooth 1996a)
- (12') B: [Even [the chairman]_{F2} is [only [considering [younger candidates]_{F1}] ~1] ~2]

In Buring's and Rooth's examples, however, both of their operators are propositional. In our examples, the narrower scope focus operator (*viz.* the adnominal ER) is sub-propositional. From a pragmatic point of view, we may nonetheless consider the associate of an adnominal ER as a second occurrence in the sense that, in order to be felicitous, the associate of an adnominal ER must be either discourse-old (cf. 13-14 cited in Creswell 2002) or uniquely identifiable (cf. 15, Siemund 2000) ¹:

- (13) a. #Our delivery boy saw **an old man himself** sitting in the lobby today.
 b. Our delivery boy saw **the old man himself** sitting in the lobby today.
- (14) a. #I bought my sandwich from **that food truck itself**. (pointing)
 b. I bought my sandwich from **the food truck itself**.

¹ Creswell also attributes the notion that adnominal ERs are incompatible with discourse-new associates to Leskosky 1972, Verheijen 1986, McKay 1991, Baker 1995 and Kemmer 1995).

- (15) *Context: The speaker and addressee are at a party where it is reasonably obvious who the guest of honor is.*
No one seems to be having fun at this party. [pointing at guest of honor]
He himself seems to be completely depressed by the look on his face.

Edmondson & Plank (1978) note that the unacceptability of examples like (13a, 14a) cannot be merely syntactic, since specific indefinites are compatible with the adnominal ER.

- (16) A: All Cretans lie.
B: Where did you hear that?
A: **A Cretan himself** told me.

Cunningham (2009) shows that the unacceptability is likewise not due to semantic type incompatibility, e.g. quantifiers (cf. 17).

- (17) Every year, the Graduate School hosts a chili cookoff between all of the different departments at Brown. It is required that some member of each department serve at least one bowl of chili to the deans. Usually the serving of chili is a job for graduate students (serving being a rather menial task for the higher ranked members of the department), but this year **every department chair himself** served some chili.

Rather, (13a-14a) are pragmatically odd. The reason is that the adnominal ER carries pragmatic import, contra the FOCUSED ASSERTION OF IDENTITY approach. The adnominal ER contributes a salient or accommodatable ordering relation *R* which ranks alternatives to the associate. As a necessary precondition, there must also be a salient or accommodatable referent of the associate. In other words, there must be a most highly ranked referent. As Creswell observes, “speakers cannot reasonably expect addressees to share beliefs about unfamiliar entities” (48).

In the rest of the chapter, I present in each section a different focus configuration and corresponding prosodic pattern. I begin with the focus semantics of a constructed example and the corresponding *f0* topline which is predicted on the extended FOCUS SENSITIVE OPERATOR account. I then illustrate the configuration, when possible, with real examples drawn from a

web-harvested corpus of speech tokens containing the string *he himself*. As described in Chapter 2, I collected a total of 232 true tokens.

I consider the following 9 focus configurations:

- clausal focus
- argument (i.e. subject) focus
- adnominal ER focus
- double focus on subject and predicate
- double focus on adnominal ER and predicate
- focus below the clause on adnominal ER
- focus below the clause on subject
- predicate focus
- overlapping associate focus

Discussion of the double focus configuration allows me to consider a third semantic approach, DOUBLE OPPOSITION, which may also be considered deterministic, although the most explicit account due to Creswell (2002) does not invoke focus explicitly.

At this point, a warning is in order. As anyone who has dealt with naturally occurring speech data knows, real speech is “messy”: for example, automatic pitch trackers fail on unvoiced segments; speakers produce utterances with unexpected pauses, hesitations and other disfluencies; speakers take for granted background information which their intended listeners know, but we as unintended listeners do not. In the absence of statistically quantifiable measurements and a complete and explicit model of the discourse, there will inevitably be room for different theoretical interpretations of individual data. However, the empirical generalization that emerges from the individual examples is, I believe, strongly robust: many more than one focus configuration is attested for the adnominal ER.

2 *Clausal focus*

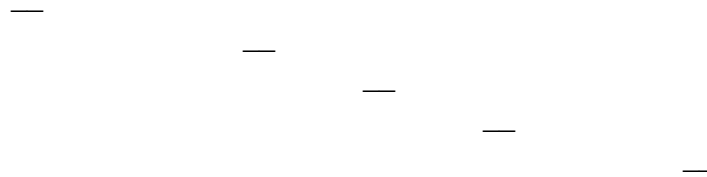
I begin with perhaps the simplest configuration: clausal or “all new” focus. It is the simplest configuration, yet perhaps the most overlooked in the literature on ERs.

I represent this configuration with a propositional focus operator for which the scope and focus associate are the entire proposition. The focus semantic value under existential closure f , that there are salient propositions of type $\langle st \rangle$, will be trivially entailed by any propositional antecedent.

- (18) a. What's new?
 b. [[[The Provost]_{F2} ~2] **himself**] **is going to chair some committee**]_{F1} ~1

$$(18') \quad f = \exists p_{\langle st \rangle}. [p]$$

The prosody of the utterance, in particular the distribution of pitch accents will be governed by independent phonological principles. Typically, all lexical words will be realized with $f0$ peaks and the height of successive pitch accents will decrease over the utterance.



- (19) [[[The Provost]_{F2} ~2] **himself**] **is going to chair some committee**]_{F1} ~1

Since the adnominal ER is not narrowly focused, STRESS-F does not predict that it will have greatest prominence within the domain corresponding to the sentence, contra the FOCUSED ASSERTION OF IDENTITY approach.

The first corpus example is a street interview with a 25-year old Pasadena City College student about his voting intention for the 2008 US Presidential election. Then-Senator Obama is highly prominent in the context and one can infer that the student ranks him highly relative to other salient individuals. There is no explicit antecedent in the discourse of the form 'x has seen it [minority groups getting along with one another] happen' which would license a narrow focus on the subject *he himself*. Nor is there an explicit relational function u relating Obama to other

individuals (e.g. Obama’s advisors). An antecedent of the form ‘*u*(Obama) has seen it happen’ would be required to license narrow focus on the adnominal ER.

While the student’s speech exhibits a limited pitch range, impressionistically, all of the content words are in fact realized with a pitch accent, such that the associate *he* has the highest *f*0 peak and subsequent accents have slightly lower *f*0 peaks. It is not the case that *himself* has the greatest intonational prominence within the utterance.

- (20) Barack Obama was talking about how there’s a misunderstanding that one minority group can’t get along with another such as African Americans and Latinos. And he’s said that

[he himself has seen it happen]

that they can and he’s been involved with groups of other minorities.

- (20’) [[[he]_{F2 ~2}] himself] has seen it happen]_{F1 ~1}

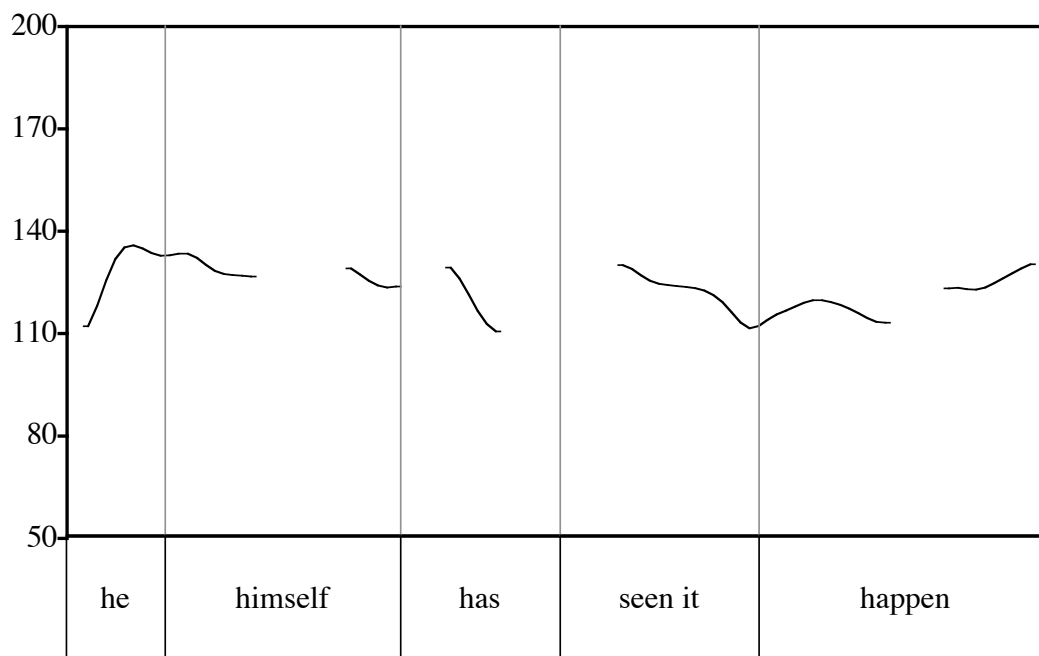


Figure 1. Smoothed *f*0 track (in Hz) of (20). Clausal focus configuration.

The second corpus example (21) is part of a discussion of 20th century Russian and Soviet poet/playwright Vladimir Mayakovsky. Mayakovsky's importance from the speaker's perspective is clear both in virtue of the artist being the subject of discussion and from the speaker's admiration for him. The Communist party and Lenin are both explicitly mentioned and are salient alternatives to Mayakovsky. There is no explicit antecedent of the form 'x was always a really valiant intractable individualist' which would license narrow focus on the subject, nor is one salient. There is also no explicit function *u* relating Mayakovsky to either Lenin or the Communist party (although one might conceivably accommodate such a relation) An antecedent of the form '*u*(Mayakovsky) was always a really valiant intractable individualist' would be required to license narrow focus on *himself*.

All of the content words have intonational prominence: the *f0* peak on *he* is highest and the heights of subsequent *f0* peaks decrease over the utterance. Narrow focus on either the subject *he himself* or the adnominal ER would require reduced *f0* on the rest of the utterance, which is not the case.

- (21) He resisted being a member of the party. But he was tucked into the machinery. He was present at Lenin's death. And he was trying to make it work. He kept believing that it was possible to reinvent society.
 [But **he himself** was always a really valiant intractable individualist]
 and he began to be criticized for that and vilified for it and increasingly became estranged from his contemporaries.

- (21') [[[he]_{F2} ~2] himself] was always a really valiant in tractable individualist]_{F1} ~1

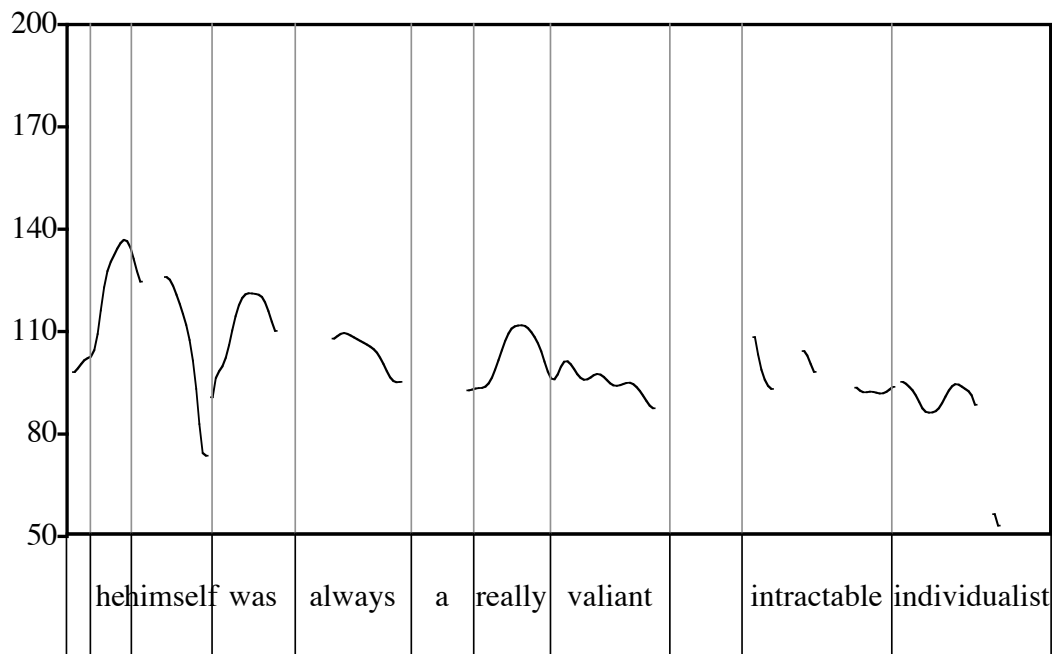


Figure 2. Smoothed f_0 track (in Hz) of (21). Clausal focus configuration.

In (22), political commentator Rachel Maddow discusses the film *Body of War* about the U.S. war in Iraq with director Phil Donahue. The associate of the adnominal ER, *the President*, is ranked against the alternative *Congress*. Although there is no explicit function relating *the President* and *Congress* which would be required to license narrow focus on *himself*, one could in principle accommodate one. With narrow focus on *himself*, however, STRESS-F would require *himself* to be realized with greatest prominence within the domain corresponding to the sentence; *the President* and following material would have reduced f_0 . This is not consistent with the f_0 track of Maddow's actual utterance, in which *he* has a higher f_0 peak than *himself*.

The predicate *determine whatever reasons* also does not have an explicit discourse antecedent which would license narrow focus on the subject; again, one could in principle accommodate one. Narrow focus on the subject would require greatest prominence on the

domain corresponding to *he himself*; *determines* would be relatively non-prominent. This is also inconsistent with Maddow's actual *f0* track.

- (22) And there is that moment when he says he could boil down this resolution. We don't need all those pages. What this resolution says is, We Congress, I'm paraphrasing here, but We Congress essentially give the President the right to do with the military whatever he wants to do to Iraq for however long he wants to [for whatever reasons **he himself** determines.] We are handing over our Constitutional responsibilities."

- (22') [[[he]_{F2 ~2}] himself] determines *t*_{F1 ~1}

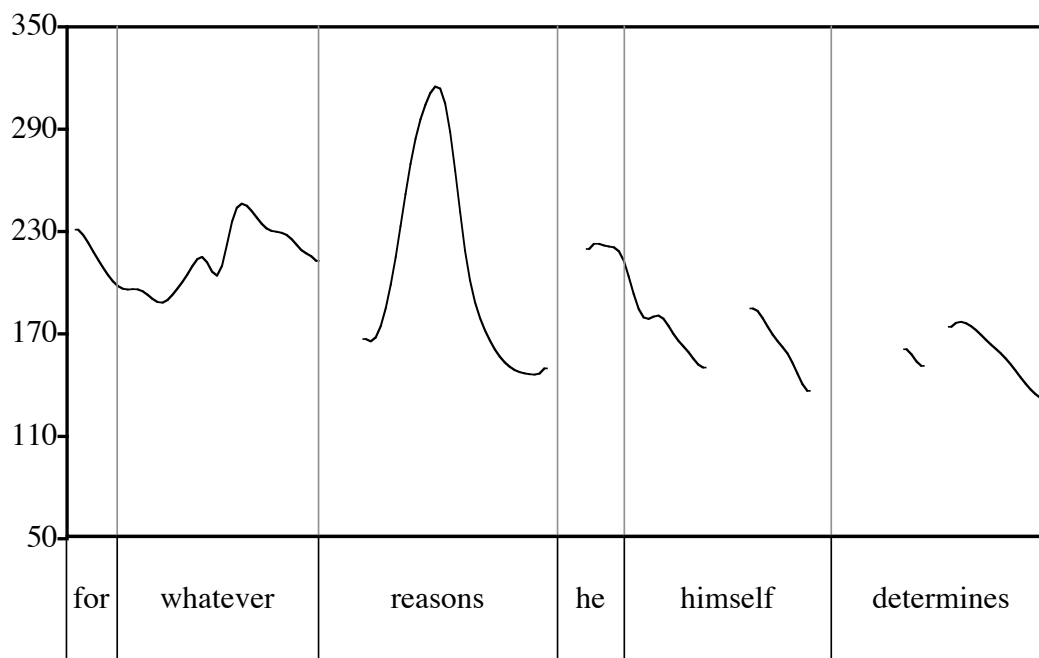


Figure 3. Smoothed *f0* track (in Hz) of (22). Clausal focus configuration.

Finally, a number of examples in the corpus are of the general form *He himself said/ admitted/realized x* with a speech act verb. Because of the additional clause and the additional phonetic material, these examples are particularly illustrative. If the adnominal ER were narrowly focused, as predicted on the FOCUSED ASSERTION OF IDENTITY approach, it must be already explicitly mentioned or salient that someone standing in a relation to the associate has

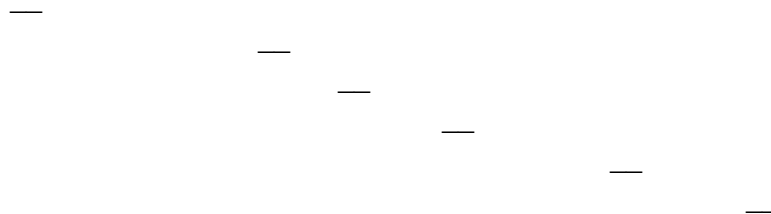
produced the reported speech or attitude. For example, in (23) there must be a salient discourse antecedent of the form ‘*u*(President) said he had committed an error’. We might infer that someone related to the Provost, or perhaps simply the speaker, said that the Provost had committed an error. Prosodically, STRESS-F requires greatest prominence on himself; *f0* on the material that follows be compressed.

- (23) a. The Provost himself said that he had committed an error.
 b. [[[the Provost]_{F2} ~2] [himself] _{F1} said that he had committed an error] ~ 1]

While narrow focus on *himself* is the only possible configuration on the FOCUSED ASSERTION OF IDENTITY, it is only one of several possible configurations, given an appropriate discourse antecedent, on the extended FOCUS SENSITIVE OPERATOR approach. Among other possibilities within the extended FOCUS SENSITIVE OPERATOR approach, (23) may also occur with causal focus (cf. 24), and corresponding prosody with descending *f0* peaks (cf. 25).

- (24) a. What’s new?
 b. [[[The Provost]_{F2} ~2] himself] said that he had committed an error]_{F1} ~1

(24') $f = \exists p_{\langle st \rangle}. [p]$



- (25) [[[The Provost]_{F2} ~2] **himself**] said that he had committed an error]_{F1} ~1

It is possible to tease apart the two configurations by making explicit that the predicate does not hold of any other individual. In (26), for example, the continuation that no one else (other than John) will admit that John is in serious financial trouble strongly disfavors an

accommodation that someone standing in a relation to John *u*(John) will admit that he is in financial trouble. Accordingly, the continuation is highly marked when *himself* is narrowly focused. With clause focus, however, the continuation is felicitous. The adnominal ER in (26b) simply signals that the speaker ranks John highly on some salient relation, and signals nothing about individuals standing in a relation to John.

(26) John is in serious financial trouble,

—
— — — — —
— — — — —

a. and **he himSELF** will admit it publicly, # although no one else will.

—
— — — — —
— — — — —

b. and **he himself** will admit it publicly, although no one else will.

Turning now to an attested corpus example, consider (27). Vermont Governor and Chairman of the Democratic National Committee Howard Dean is criticizing Republican presidential candidates and, in particular, John McCain. Host Sam Seder reports McCain's repeated admission that McCain doesn't know about economics. *A priori*, none of the possible focus configurations are impossible for Seder's utterance. It could be the Seder believes it salient that Dean or some other individual had said McCain didn't know about economics (narrow focus on the subject *he himself*); or that someone in a salient relation to McCain said that McCain didn't know about economics (narrow focus on the adnominal ER *himself*); or that someone related or not said that McCain does, in fact, know about economics (focus on the predicate or focus on the negation). Rather, McCain's admitting to not knowing about economics is presented as entirely new information in the discourse: Seder is, in my

interpretation, suggesting that McCain does not promote economic policy as a strength, such that Dean's criticism may be overly strong.

(27) Governor Dean: And he [=McCain] thinks the solution to the economic problems that we have is to cut corporate taxes. I mean, the guy is completely out of touch with where Americans are.

Interviewer: Well, Governor, to be fair

[**he himself** has said on multiple occasions he knows nothing about economics]

(27') [[[he]_{F2 ~2}] himself] has said on multiple occasions he knows nothing about economics]_{F1 ~1}

That Seder's utterance is produced with clause focus is confirmed by the prosody of (27), as represented by the f_0 track. All content words are realized with uncompressed f_0 peaks and there is a general trend of descending f_0 peaks. The f_0 peak on *himself* is not the highest within the utterance, which would be required if there were narrow focus on the adnominal ER.

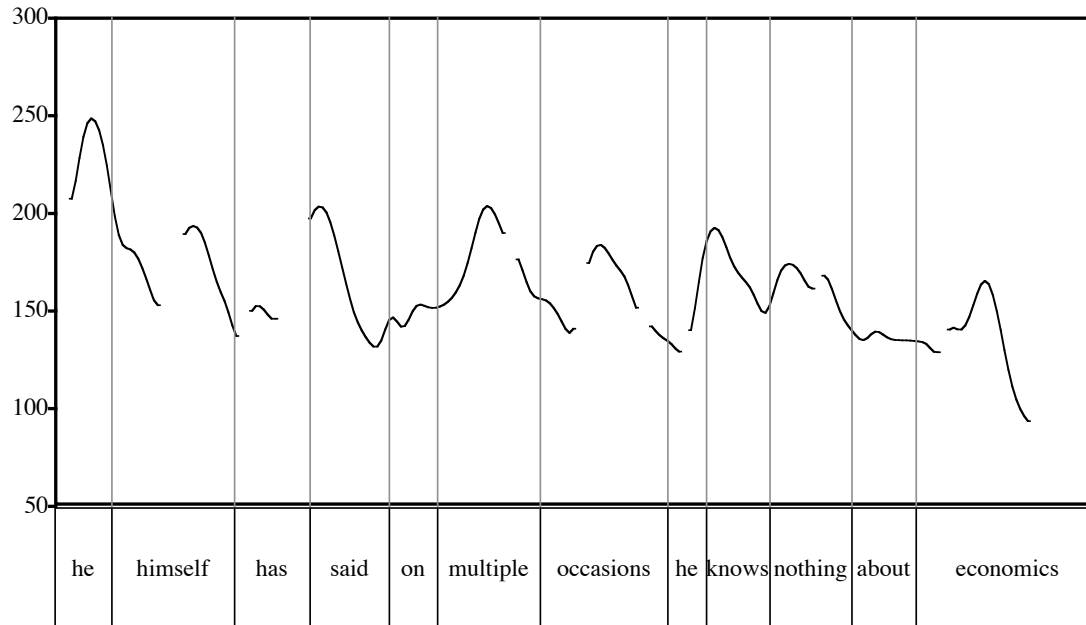


Figure 4. Smoothed f_0 track (in Hz) of (27). Clausal focus configuration.

Finally, note that the same focus configuration may be observed for other postnominal focus operators. In (28), St. Louis radio host John Carney offers his opinion on the state of golf following the news of marital infidelity by golfer Tiger Woods. The relative clause *that he has brought back to the game of golf* is presented as new information. There is no explicit or easily accommodated antecedent of the form ‘x brought back money, people and interest to the game of golf’ which would license narrow focus on the subject *he alone* or on the focus operator *alone*.

- (28) *But the real people that are affected, the sponsors and the PGA, I hadn't really thought about that... but the money and the people and the interest [that **he alone** has brought back to the game of golf] is making all of those players a lot of money.*

- (28') $[[[he]_{F2} \sim 2] \text{ alone}] \text{ has brought back } t \text{ to the game of golf}]_{F1} \sim 1$

Examining the corresponding *f0* track, we observe that all of the content words are realized with *f0* peaks; the focus associate *he* has the greatest *f0* peak and subsequent peaks decline over the utterance. This prosody is consistent with clausal focus, but not a configuration with narrow focus on the subject *he alone* or on the focus operator *alone*.

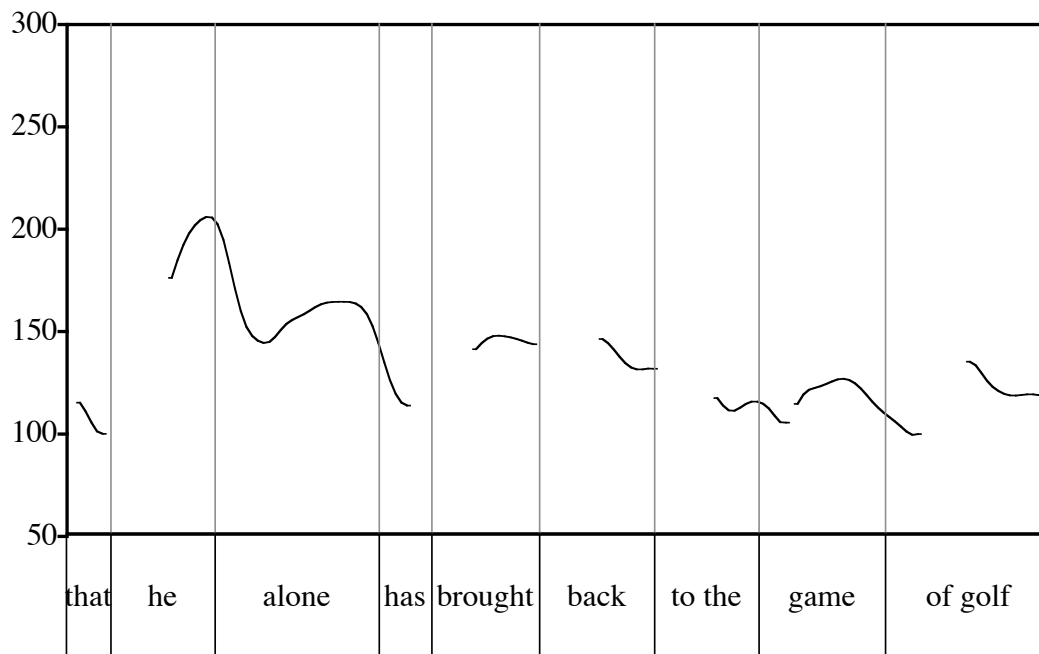


Figure 5. Smoothed f_0 track (in Hz) of (28). Clausal focus configuration.

In this section, I have demonstrated cases of clausal focus in sentences containing an adnominal ER. The clausal focus configuration is distinct from narrow focus configurations in both discourse antecedence requirements and prosodic realization. I also observed that broad clausal focus and narrow focus on the adnominal ER result in different inferences, which we can bring out with explicit continuations. And we examined several naturally-occurring examples of the clausal focus configuration.

3 *Argument focus*

In the next configuration, the constituent consisting of the adnominal ER and its associate (e.g. *he himself*) is the focus associate of a propositional focus operator. Since the adnominal ER can co-occur with an associate in any argument position (cf. Chapter 4), I refer to this as argument focus. In the corpus examples, of course, the argument is always the subject; this is an

artifact of the search methodology, since the search string *he himself* has nominative morphology.²

The subject focus configuration differs from the clausal focus configuration in that rest of the proposition—in the corpus examples, the verbal predicate—has an explicit or accommodatable antecedent. In (29b), for example, the existentially closed focus semantic value *f* is entailed by either of the antecedents in (29a). (Semantically, the meaning of the question is modeled as the set contextually constrained of possible answers.)

The subject focus configuration differs from the configuration with narrow focus on the adnominal ER in that it does not require an explicit or accommodatable relational predicate *u* to hold of the associate.

- (29) a. Who will chair the committee? *or* John will chair the committee.
 b. The Provost himself will chair the committee.
 c. [[[[The Provost]_{F2}~2] himself]_{F1} will chair the committee] ~1

- (29') $k = \forall x_e.[x \text{ will chair the committee}]$ *or* 'John will chair the committee'
 $f = \exists x_e.[x \text{ will chair the committee}]$

Prosodically, the subject is realized with greatest prominence in the domain corresponding to the scope of focus, which is clausal. Internally, the subject will have “default” prosody: both the associate and the ER will be realized with pitch accents, the second lower than the first. Pitch accents on the material following the subject, if any, will be compressed.

- (30) [[[[The Provost]_{F2}~2] himself]_{F1} will chair the committee] ~1

² The existence and acceptability of the adnominal ER with a pronoun in other positions is debatable, cf. him himself. I have nothing to contribute on the matter here.

Anecdotally, the subject focus configuration occurs less frequently in the corpus, but there are nonetheless several clear examples. In (31), the host of a forum on “Baseball as America” introduces one of several “renowned baseball experts”. With narrow focus on the subject *he himself*, (31) requires a discourse antecedent of the form ‘x has become part of baseball legend’. No such antecedent is explicitly available, but it is highly salient in the context and could easily be accommodated.

The subject focus configuration is confirmed by the *f0* track. Both the associate *he* and the adnominal ER *himself* are realized with *f0* peaks; the *f0* contour in the rest of the utterance is, in comparison, quite flat. Within the subject, the *f0* peak realized on *he* is slightly higher than the *f0* peak on *himself*. Note that a lower or absent *f0* peak on *he* would be consistent with narrow focus on the adnominal ER, rather than on the subject.

(31) *Bill James pioneered Sabermetrics, giving us many new statistics through which to look at players and teams. And in the process,*
[he himself has become part of baseball legend]

(31') [[[[he]_{F2} ~2] himself]_{F1} has become part of baseball legend] ~1
k = ‘There are people who have become part of baseball legend’
f = $\exists x_e$. [x has become part of baseball legend]

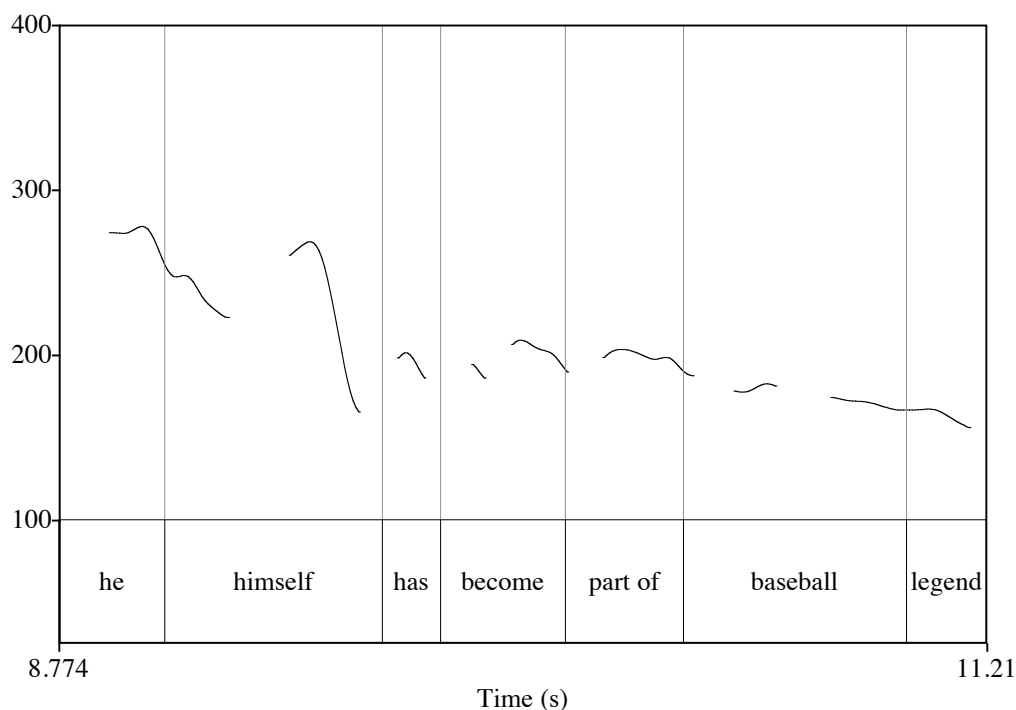


Figure 6. Smoothed f_0 track (in Hz) of (31). Argument focus configuration.

My second corpus example of the subject focus configuration, (32), is remarkable for the length of the predicate. In this NASA podcast about the history of solar flares, the speaker discusses the discovery of geomagnetic storms and their effect on telephone communication. The predicate of interest—*become aware of the effects of solar geomagnetic storms on terrestrial communications*—while rather long in duration, has an easily accommodated antecedent. In other words, it is highly salient in the context that the predicate is true of some individual, namely the scientist Louis Lanzerotti. Consequently, the predicate is realized with a noticeably reduced f_0 contour. I also provide an f_0 track for the rest of the sentence for the sake of comparison.

It is not possible to make claims about statistical significance of f_0 differences (recall that the nature of the investigation in this chapter is qualitative and impressionistic, rather than quantitative); however, note that the f_0 movement in the predicate is clearly less dynamic (all

peaks are below 130 Hz) than the f_0 movement in the *when*-clause that follows.

Impressionistically, the predicate is realized with a much flatter intonation than the subject. The subject *he himself* is clearly realized with greatest prominence in the clausal domain.

- (32) *Researchers called this a geomagnetic storm. Rapidly moving fields induced enormous electric currents that surged through telephone lines and disrupted communications. “More than thirty-five years ago, I began drawing the attention of the space physics community to the 1859 flare and its impact on telecommunications,” says Louis Lanzerotti, retired distinguished member of the technical staff at Bell Labs and current editor of the journal Space Weather. [He himself became aware of the effects of solar geomagnetic storms on terrestrial communications] when a huge solar flare on August 4th, 1972 knocked out long-distance telephooone communication across Illinois.*

- (32') [[[[he]_{F2}~2] himself]_{F1} became aware of the effects of solar geomagnetic storms on terrestrial communications] ~1
 k = ‘The space physics community became aware of the effects of solar geomagnetic storms on terrestrial communications’
 f = $\exists x_e$. [x became aware of the effects of solar geomagnetic storms on terrestrial communications]

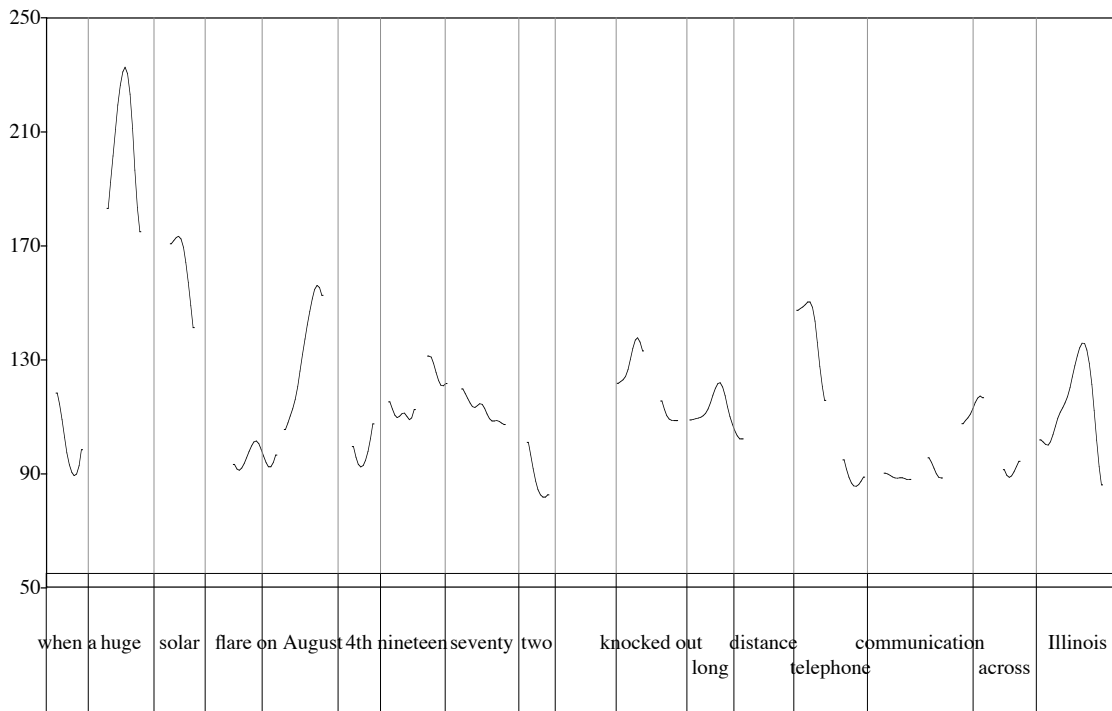
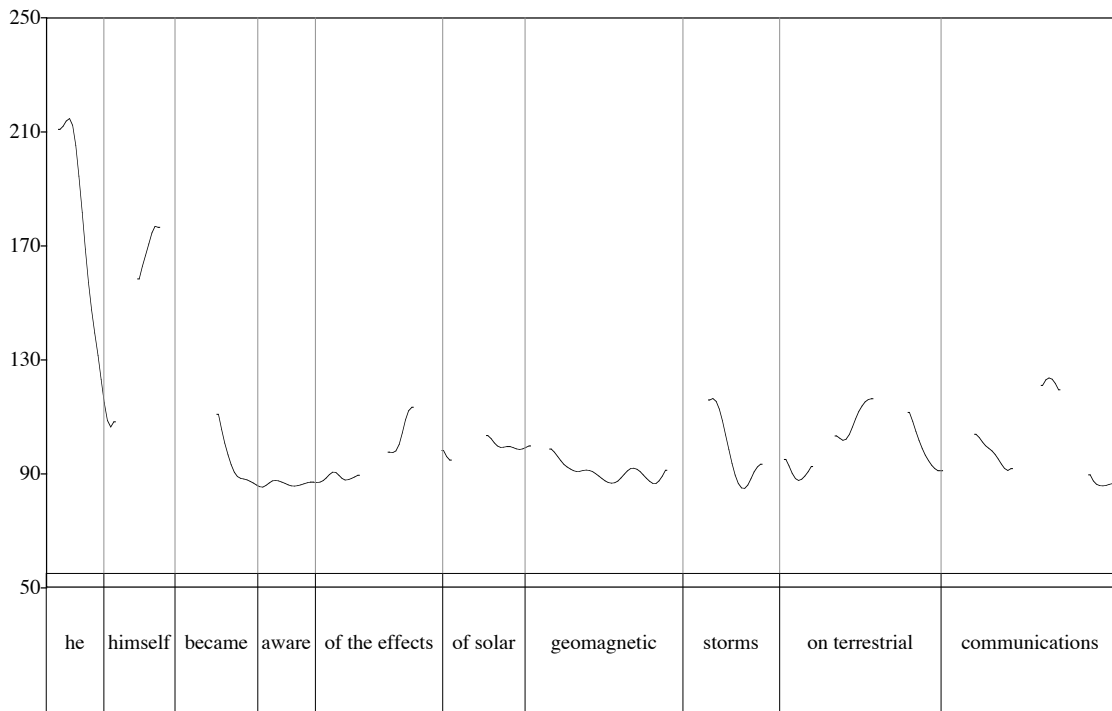


Figure 7. Smoothed f_0 track (in Hz) of (32). Argument focus configuration.

Within the subject, the f_0 peak on the associate *he* is higher than the adnominal ER *himself*. If *himself* were realized with a higher f_0 peak and *he* were realized with a lower f_0 peak, that would be consistent with a configuration of narrow focus on the adnominal ER.

In this section, I motivated a second kind of focus configuration and corresponding prosodic pattern—focus on the subject—not predicted by the FOCUSED ASSERTION OF IDENTITY approach and observed naturally occurring examples from the web corpus. In the next section, I review the configuration of focus on the adnominal ER, which is possible on both semantic approaches.

4 *Focus on the adnominal ER*

On both semantic approaches, the adnominal ER is the identity function, a semantic function mapping an individual to itself. Alternatives to the identity function will be functions mapping an individual to another, non-identical individual. Therefore, in order to license focus on the adnominal ER with clausal scope, there must be a discourse antecedent on which the same predication is made of someone in a salient relationship to the associate.

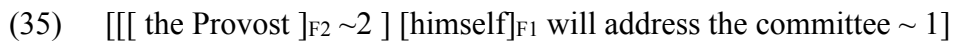
For example, in (33c) and (34c) it must be salient of someone related to the Provost that the individual will address the committee. Possible discourse antecedents with the relational predicate ASSISTANT-OF() are given in (33a) and (34a).

The adnominal ER in (33b) has an exclusive interpretation; the adnominal in (34b) has an additive interpretation.

- (33) a. The Provost's assistant will address the committee.
 b. No, the Provost himself will address the committee.
 c. [[[the Provost]_{F2} ~2] [himself]_{F1} will address the committee ~ 1]
- (33') $k = \text{'The Provost's assistant will address the committee'}$
 $f = \exists u_{ee}. [u(\text{Provost}) \text{ will address the committee}]$

- (34) a. The Provost's assistant will address committee.
b. Yes, in fact, the Provost himself will address the committee.
c. [[[the Provost]_{F2} ~2] [himself]_{F1} will address the committee ~ 1]
- (34') $k = \text{'The Provost's assistant will address the committee'}$
 $f = \exists u_{ee}. [u(\text{Provost}) \text{ will address the committee}]$

According to STRESS-F, the adnominal ER must be realized with greatest prominence within the clause; other material will have reduced prominence. In (35), *himself* will have the highest *f0* peak; *the Provost*, as prenuclear material, may have a lower *f0* peak; and the predicate, as postnuclear material, will be realized with compressed *f0* peaks, if any.



The configuration with narrow focus on the adnominal ER is similar to the previous configuration with narrow focus on the subject: the predicate is realized with reduced prominence in both. The two configurations differ in the relative prominence of the associate and the adnominal ER: the adnominal ER must be realized with greater prominence on the adnominal ER focus configuration; the associate will be realized with greater prominence on the subject focus configuration (cf. 30).

Before we begin examining naturally-occurring examples of this configuration, we also need to distinguish the focused adnominal ER from the additive ER in post-auxiliary position (cf. 36b). In Chapter 4, I alluded to the similarity between the two in the absence of an overt auxiliary.

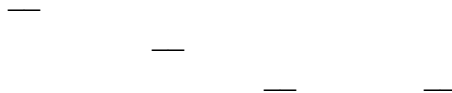
- (36) a. The Provost's assistant will address the committee.
b. Yes, in fact, the Provost **will himself** address the committee.

Like the narrowly focused adnominal ER, the additive ER also requires a discourse antecedent in which the same predication is made for some alternative; unlike the narrowly focused adnominal ER, however, the individual need not stand in some salient relation to the associate.

For example, in Chapter 4, section 4.2, we considered examples with the Pope, the President and Lady Gaga, individuals between whom it is difficult to accommodate some salient relation. We can use this property to tease apart the two kinds of ER. In (37b), the additive ER does not require a discourse antecedent with an individual in a salient relation to the Pope; Lady Gaga's addressing the committee, even though there is no salient relation relating her to the Pope, is sufficient to license (37b). In (38b), the focused adnominal ER does require an antecedent with an individual in a salient relation to the Pope; Lady Gaga's addressing the committee is therefore not sufficient to license (38b). Prosodically, (37b) has a prosody similar to the related utterance with additive *too* or *also*: cf. *the Pope will {also} address the committee*. The subject *the Pope* and the additive ER *himself* are both realized with *f0* peaks; the *f0* contour on the predicate *address the committee* is reduced. In (37), *himself* is realized with an *f0* peak; *f0* material on the rest of the utterance is reduced.

ADDITIVE ER (overt auxiliary)

- (37) a. Lady Gaga will address the committee.



- b. Yes, in fact, the Pope **will himself** address the committee.

ADNOMINAL ER (NARROW FOCUS ON THE ER, overt auxiliary)

- (38) a. Lady Gaga will address the committee.

— — — — —

- b. ? Yes, in fact, the Pope **himself will** address the committee.

When there is no overt auxiliary, utterances contrasting only in the additive ER and the focused adnominal ER will be string-ambiguous. They will nonetheless have distinct discourse requirements, and distinct prosodic patterns, as in (39-40).

ADDITIVE ER (no auxiliary)

- (39) a. Lady Gaga addressed the committee.

— — — — —

- b. Yes, in fact, the Pope **himself** addressed the committee.

ADNOMINAL ER (NARROW FOCUS ON THE ER, no auxiliary)

- (40) a. Lady Gaga addressed the committee.

— — — — —

- b. ? Yes, in fact, the Pope **himself** addressed the committee.

Now, let's review some naturally occurring examples of the focused adnominal ER configuration. In example (41) from the corpus, a political commentator discusses the tactics of U.S. presidential candidate Senator John McCain two weeks before the election. We can identify the ER *himself* as adnominal rather than additive from its occurrence before the auxiliary. The preceding context provides an easily accommodated antecedent (perhaps even an explicit one, if we analyze the preceding sentence as a case of neg-raising) *Sarah Palin not going down that course* which entails that someone standing in a salient relation to McCain, namely his running

mate, is not going down that course. The adnominal ER is consequently realized with greatest prominence within the clause, in accordance with *Stress-F*: there is a high f_0 peak on *himself* and the f_0 contour compressed on material in the rest of the clause.

- (41) *But McCain hasn't gone as far as he could've. There's also Obama's controversial pastor Reverend Wright, who said some pretty fiery sermons, one of which he said instead of 'God Bless America' it should be 'God Damn America'. So McCain's kind of backed away from that. He doesn't want Sarah Palin to go down that course. [He himself isn't going down that course] So while he's got very negative, he hasn't gone the whole way.*

- (41') [[[he]_{F2} ~2] [himself]_{F1} is not going down that course ~ 1]
 k = 'Sarah Palin is not going down that course'
 f = $\exists u_{ee}$. [u (McCain) is not going down that course]



Figure 8. Smoothed f_0 track (in Hz) of (41). Adnominal ER focus configuration.

In example (42), radio host Robin Young interviews Pulitzer Prize-winning author Richard Rousseau about his involvement in a fundraising book. He and other writers tell the true-life

stories of volunteers at a hospice in Maine. Richard is discussing how he came to write the story of friend Lee, whose wife had suffered from dementia. Lee wanted to help raise money for the hospice with a book, but didn't expect to have his own story featured. We can accommodate that someone in a salient relation to Lee, namely his friends involved in fundraising, was going to be involved in that way; this is sufficient to license narrow focus on the adnominal ER and scope of focus over the clause *he himself was going to be involved in this way*. We can also accommodate that someone in a salient relation to Lee knew that they were going to be involved in that way, which licenses narrow focus on the adnominal ER and scope of focus over the clause *Lee knew that he himself was going to be involved in this way*. Both are consistent with the speaker's *f*₀ production, in which the adnominal ER is realized with greatest prominence within the utterance.

- (42) *Young: Your friend Lee had the story of his wife.*
Rousseau: When we first started talking about it,
 [I don't think Lee knew that **he himself** was going to be involved in this way]
He didn't know how I was going to be involved, that was one thing. He just wanted me to help out if I could.

- (42') [[[he]_{F2} ~2] [himself]_{F1} was going to be involved in this way~ 1]
k = 'Lee's fellow volunteers were going to be involved in this way'
f = $\exists u_{ee}.[u(\text{Lee}) \text{ was going to be involved in this way}]$

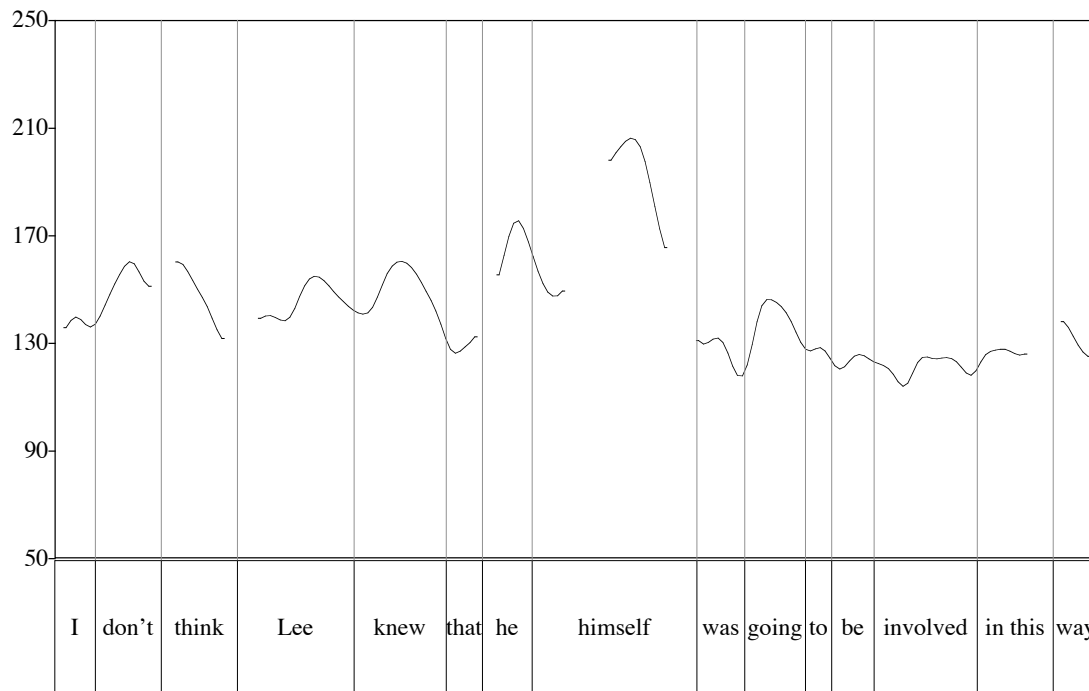


Figure 9. Smoothed f_0 track (in Hz) of (42). Adnominal ER focus configuration.

In (43), The Economist's defense correspondent Anton La Guardia discusses the performance of General Petreus in Iraq. The preceding context contains an explicit antecedent *he did better than many people thought possible* which, if we accommodate that some of those people stand in a salient relation to Petreus, entails that he did better than someone in a salient relation to him thought possible. Narrow focus on the adnominal ER is therefore licensed and the adnominal ER is realized with greatest intonational prominence within the corresponding domain. In the speaker's f_0 track, there is a high f_0 peak on *himself*; the rest of the f_0 contour is compressed.

- (43) Quite a lot of hope rests on his shoulders. He's done extremely well in Iraq, better than many people thought possible,
[better than **he himself** thought possible]

- (43') [[[he]_{F2} ~2] [himself]_{F1} thought possible he would do *x* well in Iraq ~ 1]
k = 'Petreus did better in Iraq than many people thought possible'
f = $\exists u_{ee}$. [*u*(Petreus) thought possible Petreus would do *x* well in Iraq]

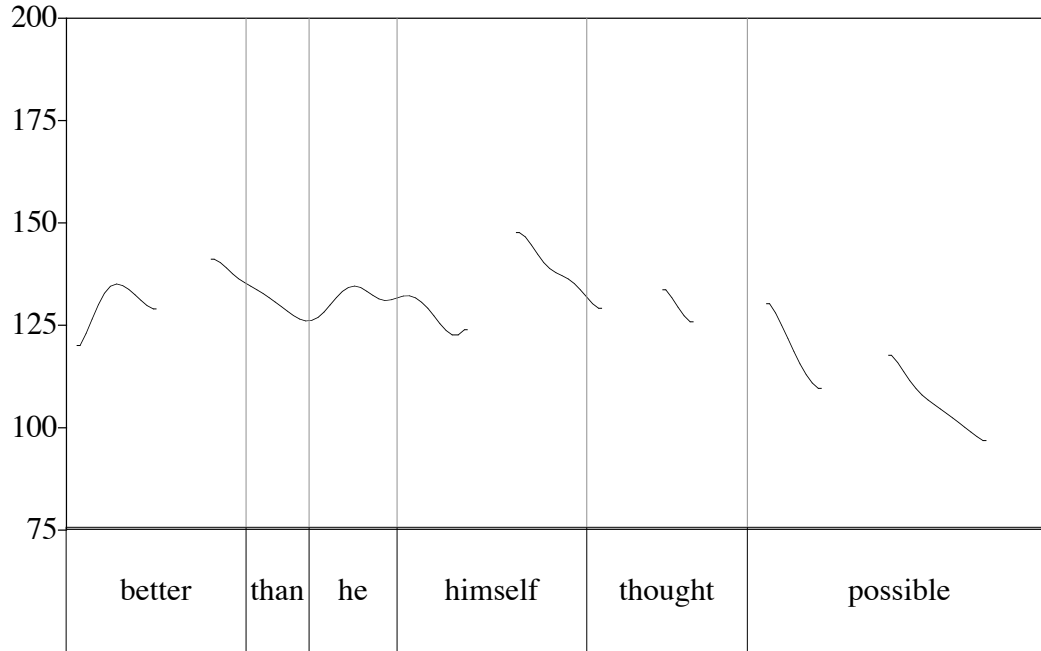


Figure 10. Smoothed *f*0 track (Hz) of (43). Adnominal ER focus configuration.

As the examples above show, the configuration of narrow focus on the adnominal ER, predicted to occur on both semantic approaches, is well attested in the corpus data. In each example, there is an explicit or easily accommodated discourse antecedent in which the same predication is made of someone standing in a salient relationship to the referent of the associate. And in each example, the adnominal ER is realized with greatest prominence—the highest *f*0 peak—within the domain corresponding to the clause. Both the discourse antecedence requirements and the prosodic realization of this configuration are distinct from the other configurations.

5 *Focus below the clause*

Thus far, we have considered configurations in which a clause-level focus operator associates with a constituent: the clause itself (Section 2), an argument containing the adnominal ER (Section 3), and the adnominal ER (Section 4). In this section, I discuss configurations in which constituents are the focused by a sub-propositional focus operator.

In the corpus examples I discuss in this section, the subject *he himself* is co-ordinated with another DP. In some cases, the referent of the DP is an individual who stands in an explicit or highly salient relation to the individual referred to by *he*, for example BUSINESS-OF(), PARTNER-OF(), ADVISORS-OF() and VICE-PRESIDENT-TO(). As we will see, association with a propositional focus operator in these examples sounds highly marked. This is because the discourse requirements are much more restrictive. The constructed example (44b), for example, requires a discourse antecedent of the form ‘*u*(the President) and his Vice Presidents should make an apology’.

- (44) a. The President’s spokesperson and his Vice Presidents should make an apology.
 b. **He himself** and **his vice presidents** should make an apology.
 c. [he [himself]_{F1} and his vice presidents should make an apology] ~1

- (44') $k = \text{‘The President’s spokesperson and his Vice Presidents should make an apology’}$
 $f = \exists_{u_{ee}}. [u(\text{the President}) \text{ and his Vice Presidents should make an apology}]$

Prosodically, *himself* will be realized with a pitch accent, and the rest of the material will be realized with pitch reduction, as represented in (45).

- — — — — —
- (45) [he [himself]_{F1} and his vice presidents should make an apology] ~1

The configuration in (44-45), while possible, is almost certainly limited to cases of correction.

For other examples with co-ordination, we require focus operators operating on constituents smaller than a clause. As discussed in Chapter 4, the alternatives semantics for focus of Rooth (1992) provides for sub-propositional focus. Consider his well-known ‘farmer’ example in (46). The two constituents *American farmer* and *Canadian farmer* serve as discourse antecedents for each other and license focus on *American* and *Canadian*, respectively. The scope of focus in both cases is the NP.

- (46) a. An **American farmer** was talking to a **Canadian farmer**.
 b. An $[[_{NP} \text{American}_F \text{ farmer}]_3 \sim 4]$ was talking to a $[_{NP} \text{Canadian}_F \text{ farmer}]_4 \sim 3]$

On a version of the FOCUSED ASSERTION OF IDENTITY approach that allows sub-propositional focus operators, narrow focus on adnominal ER may be licensed by a relational predicate that is applied to the referent of the associate, such as VICE-PRESIDENTS-TO(the President). This analysis is represented in (47b). On the extended FOCUS SENSITIVE OPERATOR analysis, the adnominal ER remains a focus sensitive operator which is focused; it is focused by a second focus sensitive operator, which is also sub-propositional (cf. 47c).

- (47) a. The President himself and his Vice Presidents
 b. $[\text{the President } [\text{himself}]_{F1}] \sim 1]$
 c. $[[\text{the Provost}]_{F2} \sim 2 \text{ } [\text{himself}]_{F1}] \sim 1]$
- (47') $k1 = \text{‘his Vice Presidents’} = \text{VICE-PRESIDENTS-TO}(\text{the President})$
 $f1 = \exists u_{\langle ee \rangle}. [u(\text{the President})]$

Prosodically, *himself* must be most prominent within the scope of focus, which is the DP containing the adnominal ER and its associate. In (47), *himself* is therefore predicted to have a higher *f0* peak than *the Provost*.

- (48) $[[\text{the Provost}]_{F2} \sim 2 \text{ } [\text{himself}]_{F1}] \sim 1]$

On the extended FOCUS SENSITIVE OPERATOR approach, focus may also be licensed on the argument, by another individual in the discourse, such as *Lady Gaga* in (49c). I use the example of *the President* and *Lady Gaga* in order to emphasize that the two individuals in this configuration need not stand in any salient relation to each other. This configuration is not possible on the FOCUSED ASSERTION OF IDENTITY approach, since on that approach the adnominal ER is meaningless unless focused.

- (49) a. The President himself and Lady Gaga
 b. [[[the Provost]_{F2} ~2 himself]_{F1}] ~1

- (49') $k1 = \text{'Lady Gaga'}$
 $f1 = \exists x_e. x$

Prosodically, the adnominal ER and its associate are focused together as a constituent, so they will have a phonologically-determined prominence relation. In (49), for example, *the Provost* and *himself* will both have high $f0$ peaks; the peak on *himself*, although not reduced, will be somewhat lower.

- (50) [[the Provost]_{F2} ~2 [himself]_{F1}] ~1

Among the naturally occurring corpus data, none support the prediction of an amended FOCUSED ASSERTION OF IDENTITY approach that the adnominal ER must be narrowly focused. They do not exhibit pitch reduction on the associate *he*, whether the adnominal is the first or second conjunct, nor do they show greatest prominence on *himself*.

The first example, (51), is taken from an interview with Marc Morano, communications director for Senator James Inhofe, who chaired the U.S. Senate Committee on Environment and Public Works. He is discussing the work of climatologist and global warming skeptic Roy Spencer. Although a relation to the associate, PARTNER-OF, is made explicit, the speaker does not

focus the adnominal ER; rather, the whole constituent *he himself* is focused and contrasted with *his partner*. Evidence of this configuration comes from the speaker's *f0* track, with declining *f0* peaks on *he* and *himself*, rather than reduced *f0* on *he* and high *f0* peak on *himself*, which we would expect if *himself* were narrowly focused.

- (51) *Let's just go back to Roy Spencer, because he's the guy you brought before Congress. He's the guy who is with the University of Alabama at Huntsville, the UAH version. And he's the guy that for ten years gave us, gave the world,*
[wrong data that **he himself and his partner** had to go back and fix]

- (51') [[[he]_{F2} ~2 himself]_{F1}] ~1 and [his partner]₁
 $k1 = \text{'Spencer's partner'}$
 $f1 = \exists x_e.[x]$

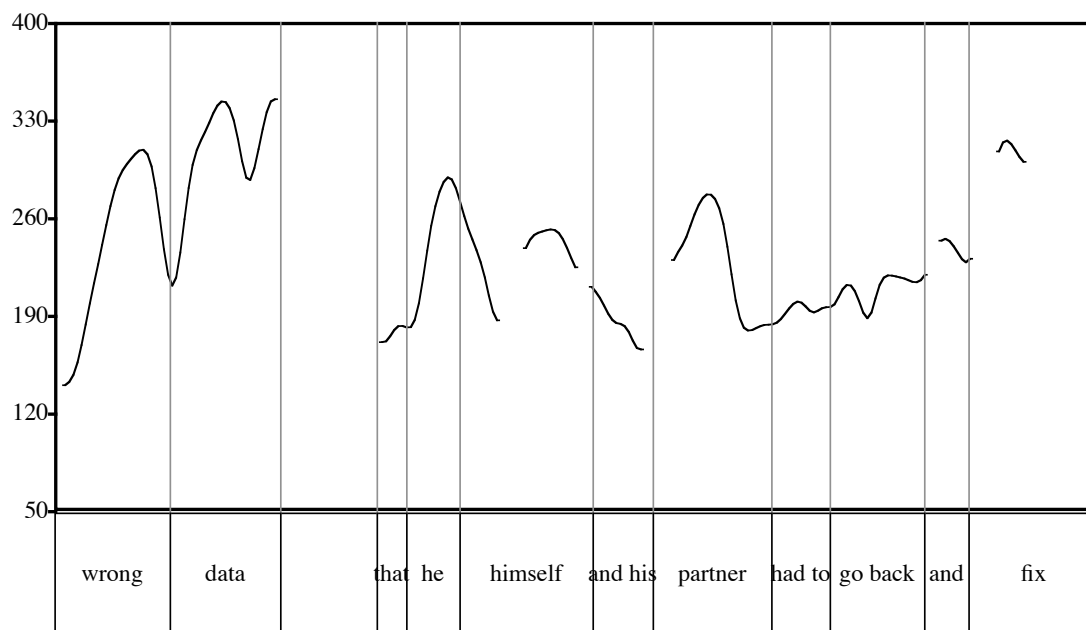


Figure 11. Smoothed *f0* track (in Hz) of (51). Sub-clausal focus configuration.

The alternative rendition with narrow focus (and greater prominence) on *himself*, I find highly marked, if not completely infelicitous. If the conjuncts are reversed (i.e. *his partner and he himself*), narrow focus seems more possible. My intuition is that this rendition requires

prosodic breaks around *he himself* indicating apposition, rather than true coordination. Further investigation with laboratory data is required; I leave this for future research.

It is clear, however, that neither order of conjuncts necessitates narrow focus on the adnominal ER. Example (52) contains a sequence in which the conjunct with the adnominal ER follows, rather than precedes its antecedent. Radio host Rachel Maddow is listing the campaign tactics of then-U.S. Senator Barack Obama. As in the previous conjunction example, the adnominal ER is not realized with a higher f_0 peak than the associate. This prosody is consistent with focus on the entire conjunct *he himself*, rather than narrow focus on *himself*, even though there is a discourse antecedent that could support narrow focus (i.e. ‘his advisors’ = ADVISORS-TO(Obama)). The observed prosody is not predicted by even an amended FOCUSED ASSERTION OF IDENTITY approach.

- (52) *I’m interested in whether or not Barack Obama is good at handling the negative stuff that’s being thrown at him. Getting Michelle Obama a very high ranking staffer this week, Stephanie Cutter, who was John Carey’s communications director; putting out the “Stop the Smears” website to try and rebut some of the lies that have been told about him;*
 [the way that **his advisors** and **he himself** pushed back so hard]
against this Giuliani-led ‘you’re soft on terrorism’ attack for the last 48 hours. Is he good at rebutting these issues?

- (52') [his advisors]₁ and [[[he]_{F2} ~2 himself]_{F1}] ~1
 $k = \text{‘Obama’s advisors’}$
 $f = \exists x_e.[x]$

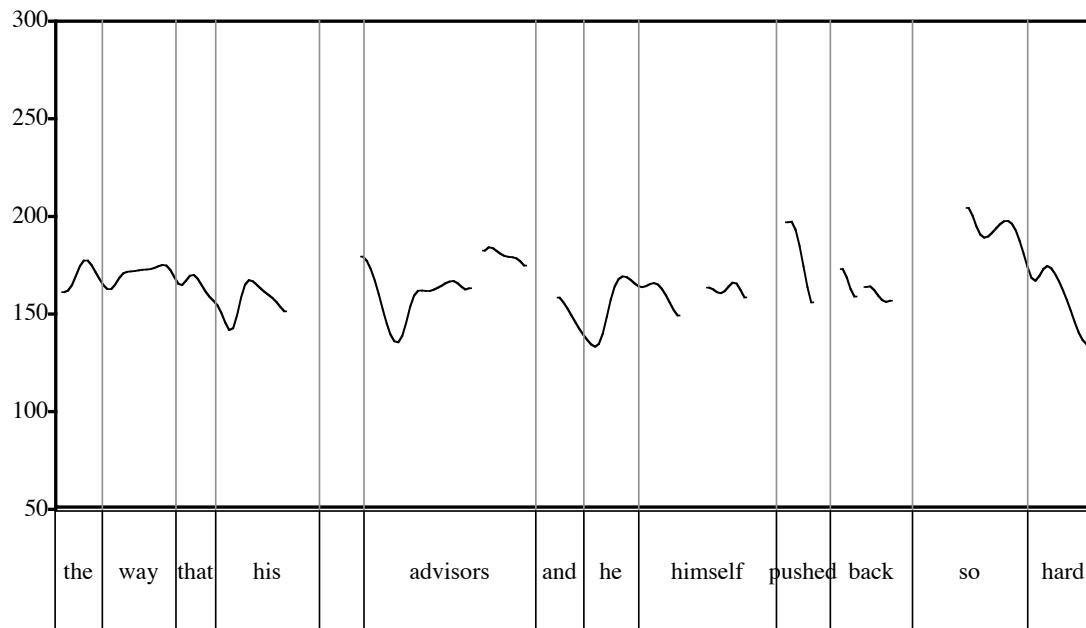


Figure 12. Smoothed f_0 track (in Hz) of (52). Sub-clausal focus configuration.

In example (52), the speaker has called in to a radio show with advice for a couple of people who have been victimized by a building repair contractor. The speaker makes a distinction between the contractor ('the guy') *qua* individual and *qua* business, explaining that the contractor could declare bankruptcy in either capacity to avoid having to pay the homeowners. This example, too, would seem to be an excellent context for narrow focus on *himself*, since there is an antecedent of the form *u(the guy)*, namely *him as a business*. The corresponding prosody for that configuration would be such that *himself* has greater prominence than *he*, and I have the intuition that a rendition with this prosody would be acceptable. The speaker's actual rendition, however, is with greater prominence on *he*, corresponding to focus on the constituent *he himself*. This prosody not predicted on the FOCUSED ASSERTION OF IDENTITY approach.

- (53) *My comment was, I actually work for a bankruptcy attorney, so I wanted to kind of reiterate what someone had already mentioned about going after the guy's license because, to just file suit against him*
[he himself or him as a business could always file bankruptcy]
and never see any money from that man.

- (53') $[[[\text{he}]_{F2} \sim 2 \text{ himself}]_{F1}] \sim 1 \text{ or } [\text{him as a business}]_1$
 $k = \text{'him as a business'} = \text{BUSINESS-OF}(\text{the guy})$
 $f = \exists_{x_e} [x]$



Figure 13. Smoothed f_0 track (in Hz) of (53). Sub-clausal focus configuration.

In the last example, the antecedent does not contain an explicit relation u , but such a relation is nonetheless easily accommodated. The extract in (54) comes from a Cambridge forum “Exploring How to Go About Recovering American Ideals” hosted by Boston College Bossey Centre Director Alan Wolfe. In addition to explicit mention of Vice President Cheney in the conjunct, Cheney is also discussed by the speaker in the preceding discourse. Nonetheless, despite the salient alternative function VICE-PRESIDENT-TO, the adnominal ER is not narrowly focused. *Himself* does not have the highest f_0 peak within its focus domain. Rather, the f_0

contour is consistent with broad focus on the subject *he himself*: a high *f0* peak on *he* and a lower *f0* peak on *himself*.

(54) Bush is not a party builder. I think you’re absolutely right that he fits the script so well for blending these Republican themes, but
[he himself and Vice President Cheney]
have never really put building the Republican party first

(54') [[[he]_{F2} ~2 himself]_{F1}] ~1 and [Vice President Cheney]₁
k = ‘Vice President Cheney’
f = ∃_{x_e}.[*x*]

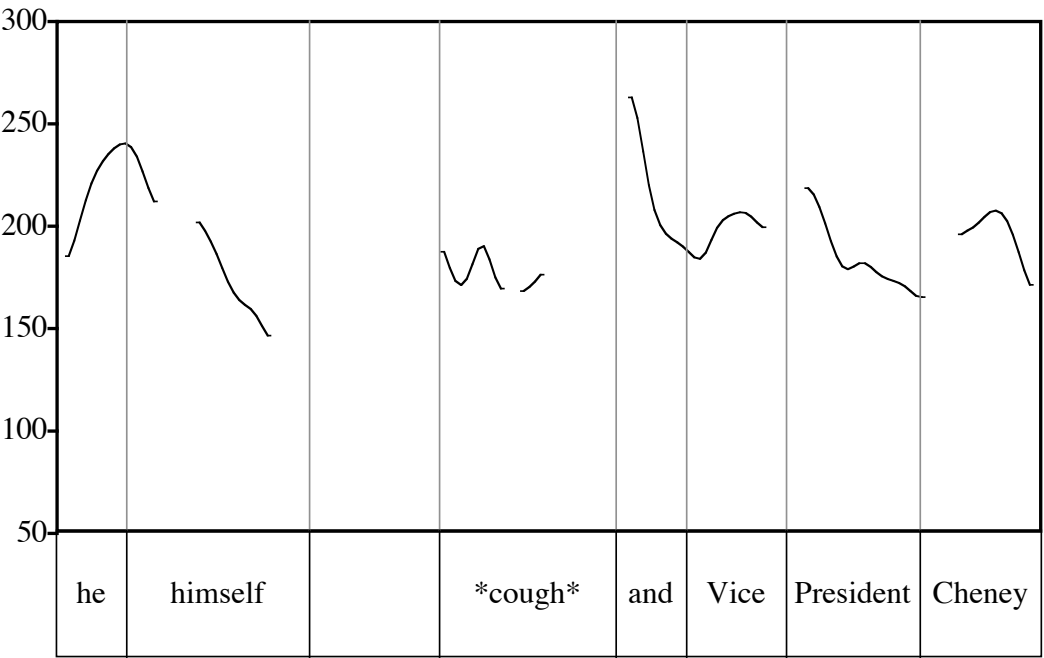


Figure 14. Smoothed *f0* track (in Hz) of (52). Sub-clausal focus configuration.

All of these corpus examples confirm that focus on the argument *he himself* in coordinate structures, rather than on the adnominal ER *himself*, is an attested focus configuration. This finding does not support the FOCUSED ASSERTION OF IDENTITY approach, which predicts only narrow focus on the adnominal ER by a propositional operator. Narrow focus on the adnominal

ER by a sub-propositional focus may or may not be possible; further experimental investigation with elicited data is required, which I leave for future research.

6 *Double focus*

We have now reviewed the most common focus configurations having a single focus associate. In this section, I examine cases of more than one focus associate. In addition to providing evidence against the FOCUSED ASSERTION OF IDENTITY approach, the double focus configuration also informs what may be considered a third approach to the semantics and pragmatics of the adnominal ER. Following Dirven (1973), I will refer to this third approach as the DOUBLE OPPOSITION approach. Dirven proposed that adnominal ERs “necessitate the presence of a double opposition” (287, cited in Verheijen 1983:277). Creswell (2002), working in the context of natural language generation, develops a related analysis that requires both that the associate be most highly ranked according among salient alternatives (in much the same way as the extended FOCUS SENSITIVE OPERATOR approach) and that a contrasting predication hold for one of these alternatives:

- (55) Speakers use [the adnominal ER] felicitously only when their addressee can infer from the context that:
- (i) the referent is the most prominent of a set of related entities
 - (ii) the predication they are making about this prominent entity contrasts with some other salient predication in the discourse
- Creswell (2002:33)

In examples (56-59) from Creswell, there is both an explicit alternative individual y to the associate of the adnominal ER x and an explicit alternative predicate Q to the predicate P that holds of the associate. (Creswell’s text corpus of 80 tokens comes from home repair manuals, agricultural-related web site, news articles and novels.)

- (56) It's my lunchtime reading, so progress is slow, but it's something to look forward to everyday (**lunch itself** isn't).
P = 'is something to look forward to'
x = 'it'
Q = 'isn't [something to look forward to]'
Y = 'lunch'
- (57) As chairman, Dan Evans cultivates a folksy feeling in the campus-like, red-brick headquarters building, where male executives are required to wear old-fashioned string ties and where **Dan himself** wears cowboy boots.
P = 'wear old-fashioned string ties'
x = 'male executives'
Q = 'wear cowboy boots'
y = 'Dan'
- (58) Instead of the wood itself splitting, the glued joints may open up.
P = 'split'
x = 'the wood'
Q = 'open up'
y = 'glued joints'
- (59) We climbed the four granite steps before the Research Laboratory. **The building itself** was of unadorned brick and rose six stories. We passed between two heavily-armed guards at the entrance.
P = 'Granite'
x = 'the steps'
Q = 'Brick'
y = 'the building'

Analyzed within the extended FOCUS SENSITIVE OPERATOR approach, (56-59) are examples of double clause-level focus. The adnominal ER is a focus sensitive operator, independently of the propositional level focus; it is this sub-propositional operator which presupposes that the associate is mostly highly ranked on some salient ordering (cf. Creswell's first property 55i).

There is also a proposition-level focus operator with two focus associates (cf. Creswell's second property 55ii). In the web-harvested speech corpus examples, the first of the two focus associates is the constituent [*he himself*] or the adnominal ER *himself*; the second of the two focus associates is the *VP* predicate, or some subconstituent of the *VP* predicate.

For example, in both (60) and (61), the predicate *meet with stakeholder groups* contrasts with *chair the committee*. In (60), there is a second focus associate *the Provost himself* which contrasts with *the company's representative*; in (61) the second focus associate is *himself* and it contrasts with *'s assistant*.

The two examples (60,61) require two slightly different antecedents. The proposition-level focus in (60b) requires an antecedent with an alternative subject-predicate pair $\langle y, Q \rangle$. For example, (60a) entails that there is some salient individual x other than the Provost and some salient predication Q other than chairing the committee such that x will Q . (61b) is slightly more restrictive in requiring an antecedent with an alternative pair consisting of a function u which holds of the Provost and a predicate Q which holds of $u(\text{Provost})$. For example, (61a) entails that there is some relation u other than identity and some predication Q other than chairing the committee such that $u(\text{Provost})$ will Q .

- (60) DOUBLE FOCUS: ARGUMENT ASSOCIATE, PREDICATE ASSOCIATE
- The company's representative will meet with stakeholder groups,
 - and **the Provost himself** will chair the committee.
 - [[[[the Provost]_{F1} ~1] himself]_{F2} [will address the committee]_{F2}] ~ 2]

(60') $k = \text{'The company's representative will meet with stakeholder groups'}$
 $f = \exists x_{ee} \exists Q_{et} . [y \text{ will do } Q]$

- (61) DOUBLE FOCUS: ADNOMINAL ER ASSOCIATE, PREDICATE ASSOCIATE
- The Provost's assistant will meet with stakeholder groups
 - and **the Provost himself** will chair the committee.
 - [[[[the Provost]_{F1} ~1] [himself]_{F2} will [chair the committee]_{F2}] ~ 2]

(61') $k = \text{'The Provost's assistant will meet with stakeholder groups'}$
 $f = \exists u_{ee} \exists Q_{et} . [u(\text{Provost}) \text{ will do } Q]$

These configurations are analogous to the single focus configurations of argument focus (Section 3) and adnominal ER focus (Section 4), respectively. In other words, the double focus configuration simply adds an additional focus on the predicate, or within the predicate. The same

arguments that I made about the single focus configurations may be carried to the double focus configurations. The double focus configuration with an argument focus associate (cf. 60) is predicted on the FOCUS SENSITIVE OPERATOR approach, and not predicted on the FOCUSED ASSERTION OF IDENTITY approach. The double focus configuration with an adnominal ER focus associate (cf. 61) is predicted on both approaches.

I will delay discussion of the prosodic realization of these configurations until Section 6.3. Having shown how one represents the double focus examples within the extended FOCUS SENSITIVE OPERATOR approach, I will discuss the implementation of the DOUBLE OPPOSITION approach and how it fails to account for the full range of adnominal ER examples.

6.1 *Double opposition approach*

The DOUBLE OPPOSITION approach (Creswell 2002) takes the double focus configuration as the default case. How are other configurations treated on this approach? The argument focus and adnominal ER focus configurations, which require an antecedent with the same predicate, Creswell also subsumes under the category of double focus. Unfortunately, she must weaken her condition (55ii) from “contrasting” predication to “related” predication, which includes non-unique alternative predicates. “The most basic type of related predication,” writes Creswell, “is simply when the same predicate P holds of the prominent entity and some member(s) of its containing set” (37).

In (62), Creswell has in mind an interpretation on which we accommodate (reasonably so in a discussion about recycling) that there are non-recycled entities which are consigned to the scrap head.³ On this interpretation, the predicate holds of both alternatives: the program and non-recycled entities. I indicate both contrasting constituents ($P/Q, x/y$) in the style of Creswell, as well as discourse antecedent k and existentially quantified focus semantic value f for her examples.

³ Other interpretations are possible, including a true double focus configuration, but we’ll see that they have different prosodic realizations.

- (62) Urbana was one of the first cities in Illinois to adopt a recycling program. Now **the program itself** may be consigned to the scrap heap.

$P = Q =$ 'consigned to scrap heap'

$x =$ 'the program'

$y =$ 'non-recycled entities'

- (62') $k =$ 'Non-recycled entities are consigned to the scrap heap'

$f = \exists x_e$ such that 'x is consigned to the scrap heap'

In addition to (62), Creswell presents two other examples of contrasting non-unique predications. In both of these, the associate and adnominal ER are conjoined with an explicit alternative. These, we might analyze as double focus of an argument and VP predicate (cf. 60); alternatively, we might analyze them as focus below the clause (i.e. constituent focus, Section 5).

- (63) Epoxy adhesives are sold in two tubes, **the glue itself** and a hardener.

$P =$ 'sold in tube'

$x =$ 'the glue'

$y =$ 'a hardener'

- (63') *argument focus*

$k =$ 'the hardener is sold in a tube'

$f = \exists x_e.[x \text{ is sold in a tube}]$

- (63'') *constituent focus*

$k =$ 'the hardener'

$f = \exists x_e.[x]$

- (64) But these may not be the safest places either for the sake of our children and pets, or for **the product itself**.

$P =$ 'these are not safest place for the sake of'

$x =$ 'the product'

$y =$ 'our children and pets'

- (64') *argument focus*

$k =$ 'these may not be the safest places for our children and pets'

$f = \exists x_e.[x \text{ may not be the safest places for something}]$

- (64'') *constituent focus*

$k =$ 'our children and pets'

$f = \exists x_e.[x]$

Although Creswell must weaken the notion of contrast, she remains committed to the DOUBLE OPPOSITION approach because, she argues, prominence alone (cf. 55i) is not sufficient to license the adnominal ER; the adnominal ER is infelicitous unless a “contrasting” predication holds of an alternative. As evidence, she offers (65), an example modified from an attested corpus example. The adnominal ER is infelicitous in this example, according to Creswell, because there is no other alternative in the discourse for which the predicate ‘needs to be refilled’ or the predicate ‘not refilled’ holds. Even though the toilet tank is “highly salient” and the “most prominent member of the set of parts of a toilet”, this prominence alone is not sufficient to license the adnominal ER.

- (65) When cold water enters a toilet and meets the warm air of the bathroom, the tank or bowl may sweat. To reduce sweating, empty the tank, dry it thoroughly, and line it with ½-inch polystyrene or foam rubber. Use a kit or cut your own liners to fit the tank walls and floor. Glue the liners in place with silicone cement, and let the adhesive dry thoroughly (at least 24 hours) before refilling **the tank (#itself)**.

I must agree with Creswell that, in a discussion of how to stop a toilet tank from sweating, the tank certainly is highly salient and also prominent. On the extended FOCUS SENSITIVE OPERATOR approach defended here, there must also be a salient ordering relation which ranks the associate most highly. The tank is in some sense important, but I find it difficult to articulate—in the context of this discussion, or more generally, the sense in which the tank might be *more important* than the (entire) toilet, the bowl, the liner or the sweat. Certainly, the predicate (i.e. *(not) refilling*) does not provide any salient ordering relation that would rank the tank most highly. The problem with this example is that nothing else in the context does either.

Creswell observes that other examples may be saved “when the distinguished member [associate] is somewhat important in almost any context”. In what I have called “big shot” examples, it is sometimes difficult to retrieve an ordering relation directly from the predicate. It is rather easy, on the other hand, to accommodate an ordering relation based on sociological

importance—at least in the case of human or human-like referents. In (66), one might accommodate a salient relation which ranks the personification of fate topmost (e.g. R = powerful beings) and, in (67), a salient relation which ranks the devil topmost (e.g. R = evilness). These examples are problematic for Creswell because of the other requirement that the adnominal ER be licensed by a salient contrasting predicate. It is not obvious, however, what the appropriate contrast to ‘had sent Robert Grant Burns into their lives’ or ‘must have defecated you into my classroom’ might be. As Creswell admits, “the salience of the related predication about other members of the set may be quite weak.”

- (66) Lite said that [Burns had no business being there] because he [=Lite] was not given the power to peer into the future and so could not know that **Fate herself** had sent Robert Grant Burns into their lives (Golde 1999, cited in Creswell 2002)

P = ‘sent Burns into their lives’

x = ‘Fate’

Q = ?

y = ?

- (67) Cried one professor after a few months of Student Schiele’s tantrums and rebellion: “**The devil himself** must have defecated you into my classroom!” (from Brown Corpus in Golde 1999, cited in Creswell 2002)

P = ‘defecated schiele into classroom’

x = ‘the devil’

Q = ?

y ?

Since the modified FOCUS SENSITIVE OPERATOR account does not require a related or contrastive predication, such examples are unproblematic. Furthermore, Creswell does not specify a linguistic mechanism, whether syntactic, semantic or pragmatic, by which the adnominal ER would contribute the felicity conditions in (55). On the modified FOCUS SENSITIVE OPERATOR account, the primary contribution of the adnominal ER is the ordering relation; a contrasting predication, if any, is signaled independently by focus at the proposition-level. Although maintaining the DOUBLE OPPOSITION approach based on the data available to her,

Creswell also anticipates this possibility of focus operating independently, speculating that contrastive pitch accent may, in fact, play a role:

Whether such a usage [of pitch accent] is informationally redundant or whether emphatic reflexives in actuality contribute only a ‘meaning’ of prominence (and the contrastiveness assumed here is contributed by the phonological focus) remains an open question. Because the tokens in the present corpus are all from written texts, an empirically-based characterization of the correlation between contrastive accent placement and emphatic reflexives is a topic for future research. (Creswell 2002:41)

In this study, we have access not only to the transcriptions of naturally occurring data, but the acoustics, as well. In the next subsection, I examine examples of double focus prosody in utterances containing the adnominal ER. The double focus configuration, just one of several possible focus configurations, is responsible for the “double opposition” meaning when it occurs and not the adnominal ER.

6.2 Prosody of the double focus configuration

The prosody of double focus has received less attention in the literature, so I will devote more attention here to the theoretical predictions and how the prosody of double focus is expected to differ from other focus configurations. Eady *et al.* (1986) test several focus configurations in the laboratory, directly comparing the double focus configuration to three other focus configurations in minimally contrastive utterances. They elicit different focus configurations on the same string by asking different *wh*-questions (cf. 68).

Their “neutral” focus condition corresponds to what I have called clausal focus and “all new” prosody. “Initial focus” corresponds to narrow focus on the subject (cf. Section 3), although in Eady *et al.*’s data the focus associate in this configuration was a single word; in our corpus dataset, the subject contains the two words *he himself*. The adnominal ER focus configuration does not correspond to any of Eady *et al.*’s conditions, since the adnominal ER is not initial. We have not reviewed any corpus examples of final single focus, although I consider

some possible candidates of “predicate focus” in Section 8. Finally, the dual focus condition is double focus on the subject and clause-final argument.

(68) Target sentence: *Don shot the puck to Kent.*

Focus conditions:

- | | | |
|----|---|----------------------------------|
| a. | BROAD / ALL NEW / “NEUTRAL” | ‘What happened?’ |
| | [[Don shot the puck to Kent] _F] ~ | |
| b. | INITIAL SINGLE FOCUS | ‘Who shot the puck to Kent’ |
| | [[Don] _F shot the puck to Kent] ~ | |
| c. | FINAL SINGLE FOCUS | ‘Who did Don shoot the puck to?’ |
| | [[Don shot the puck to [Kent] _F] ~ | |
| d. | DUAL FOCUS | ‘Who shot the puck to whom?’ |
| | [[[Don] _F shot the puck to [Kent] _F] ~ | |

As discussed in Section 2, in the broad or clausal focus configuration, the clause is both the focus domain and focus associate; therefore, phonological principles largely determine pitch accent placement within the domain corresponding to the clause. Eady *et al.* find that high pitch accents in a sequence are subject to gradual downstepping. The keywords in Figure 1 are the subject (e.g. *Don*) the first object (e.g. *puck*) and the second object (e.g. *Kent*). In the clausal focus configuration the second keyword is realized with a lower *f0* peak than the first keyword, and the third keyword is realized with a lower *f0* peak than the second. Their results confirm the predictions for *f0* peaks I made in Section 1.

As discussed in Section 3, in the narrow subject focus configuration (Eady *et al.*’s initial focus condition), the subject is realized with a nuclear accent—the last pitch accent in a phonological phrase—and subsequent material is deaccented. The *f0* contour is characterized by a *f0* peak followed by a fall to a lower *f0*. In Eady *et al.*’s data, there is a large decline from keyword 1 *f0* peak to keyword 2 *f0* peak. Keyword 3 is realized with a slightly lower *f0* peak than the *f0* peak on keyword 2.

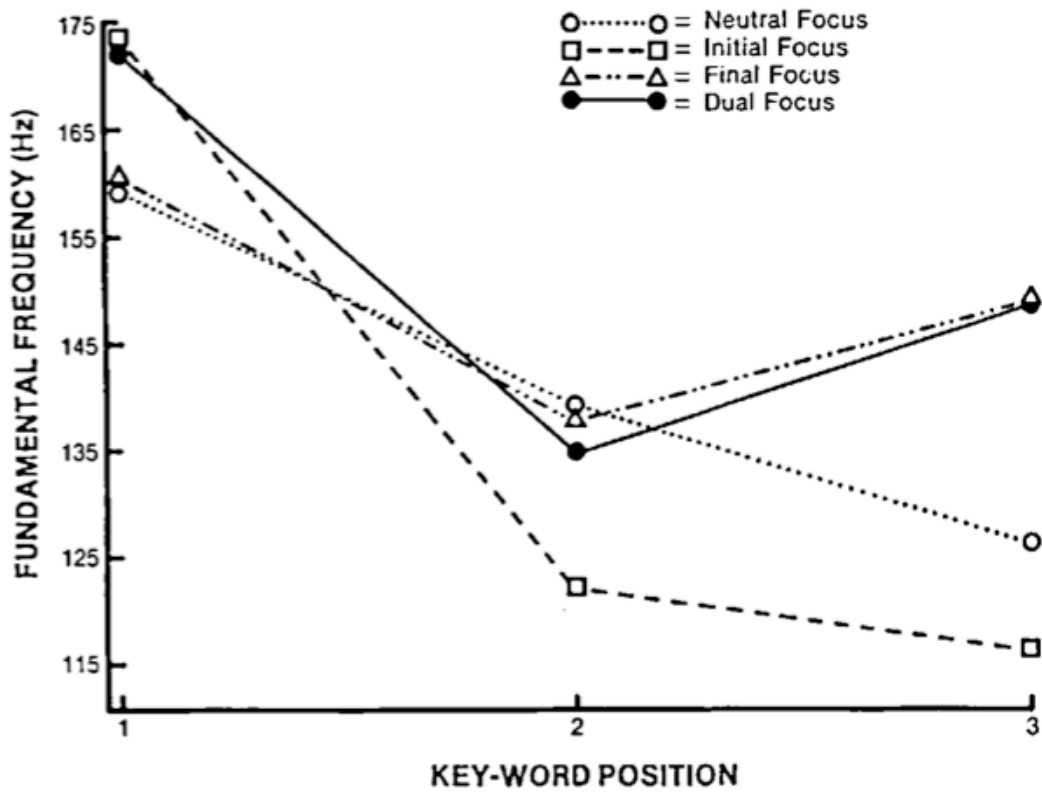


Figure 15. Average f_0 toplines for four different focus conditions (Eady *et al.* 1986:245)

In the final focus condition, Eady *et al.* found that the f_0 peak on the nuclear accent was higher than on the preceding f_0 peak, although lower than the first f_0 peak. These absolute values are not consistent with *Stress-F* since the late focus is not realized with the highest absolute f_0 value. However, compared against a baseline of sequentially downstepped high pitch accents in the clausal focus configuration, one might imagine that the f_0 peak on a late focus is perceptually most prominent.

Finally, in the double focus configuration, the f_0 peak contour has the same general pattern as late focus, with the difference that the f_0 peak realized on the first focus is much higher than it would be on the same word in the late focus condition.

Conducting an informal, blind perception test, one of the experimenters was able to identify the correct focus condition (i.e. “neutral” focus, initial focus, final focus or dual focus) at

a 96% success rate. This suggests there must be reliable acoustic cues distinguishing the different configurations.

In my representations of prominence in this chapter, I have thus far relied on *f0* peaks alone, and I have made the oversimplifying assumption that all pitch accents are high. The similarity of *f0* peaks between the double focus and final focus configurations suggest additional representation is required. On a Beckman and Pierrehumbert (1986) model of intonation, a phonological utterance contains one or more intermediate phrases, and a nuclear accent is the last pitch accent in a given phrase. According to Beckman (1996), it follows then that “to produce an utterance with double foci explicitly on two different [constituents] [...] it is necessary to make two intermediate phrases” (35). An intermediate or phonological phrase is realized with a boundary tone and pause, the latter realized by different phonetic cues such as silence or lengthening).

Welby (2003) observed that the text in (69) with a clearly double focus configuration was realized consistently in a radio speech corpus with a double nuclear accent pattern. Although the specific types of pitch accent differed, five out of six speakers reading the same string in (69), realized *functions* with a nuclear pitch accent and one of the words in *electronic probation officer* with a nuclear accent.

- (69) H* L+H* L- L%
- It [looks]_F like [a watch]_F
- H* L- H* L* L-L%
- It [functions]_F like [an electronic probation officer]_F

The apparent ambiguity between a rendition with two nuclear accents and a rendition with a late nuclear accent may also be due to a complex pitch accent L+H* on the second nuclear accent. Jannedy (2002) found that speakers consistently distinguished a double nuclear accent contour (H* L- H* L-L%) from a late peak accent contour (H* L+H* L-L%), identifying the double nuclear accent contour with a double focus configuration and the late peak accent contour

with a late single focus configuration. Phonologically, the two renditions differ only in the alignment of the low tone: in the late nuclear accent contour, the low tone is part of a pitch accent (cf. 70a); in the double nuclear accent contour, the low tone is a boundary tone (cf. 70b).

- (70) a. LATE SINGLE FOCUS
What did Ben hear?

H* L+H* L-L%
[[Ben heard [shots]_F] ~

- b. DOUBLE FOCUS
'Who heard what?'

H*-L H* L-L%
[[[Ben]_F heard [shots]_F] ~

Returning to the adnominal ER corpus data, this suggests that we will not be able to reliably distinguish the double focus configuration from the late single focus configuration from examination of the *f0* peak contour alone. Fortunately, we can in most cases rule out the late single focus configuration on semantic grounds, within any of the approaches discussed. On the FOCUSED ASSERTION OF IDENTITY approach, the adnominal ER is predicted to always be the associate of a focus operator. A late single focus by definition cannot also have an early focus. Therefore, the late single focus configuration is not predicted. On the DOUBLE OPPOSITION approach, the adnominal ER is only ever realized as part of double focus configuration. The extended FOCUS SENSITIVE OPERATOR approach *does* allow for a late single focus configuration, but requires the configuration to be licensed by an appropriate antecedent. For example, (71b) is licensed by (71a), which has the form 'he himself will P the committee'. I found only one possible example of this configuration in the web-harvested corpus, which I discuss in Section 7.

- (71) a. The Provost himself will chair the committee.
b. No, he himself will chair the committee.
c. [[[he]_{F2} ~2] himself] will [**chair**]_{F1} the committee] ~ 1

- (71') $k = \text{'The Provost himself will chair the committee'}$
 $a = \exists P_{\text{et.}} [\text{The Provost will } P \text{ the committee}]$

The results of Eady *et al.* (1986), Jannedy (2002) and Welby (2003) do suggest that the double focus configuration should have a distinct realization from the argument focus and adnominal ER focus configurations, which are both cases of early focus: they are realized with a single, early nuclear accent followed by deaccented material, in a single intermediate phrase. The interface principle *Stress-F* is satisfied in these single focus configurations because the associate—either the argument *he himself* or the adnominal ER *himself*—is realized with greatest prominence within the domain corresponding to the scope of focus, the clause.

The results also confirm that the double focus configuration is realized distinctly from the clausal focus configuration. The clausal focus configuration is realized with a series of high tones, typically downstepped, within a single intermediate phrase.

6.3 *Web examples of the double focus configuration*

Many examples of the double focus configuration are attested in the web-harvested corpus, both examples in which the adnominal ER is the first focus associate and examples in which the subject is the first focus associate. I'll begin with the former.

In (72), a reporter describes a Manhattan crane collapse. The foreman on site escaped, while some of his workers were injured or killed. The target sentence *he himself escaped the crane collapse* is licensed by the salient discourse antecedent *A worker of Blair's was injured in the crane collapse*, which contains an alternative relational predicate-verbal predicate pair: <WORKER-OF(), was-injured-in()>.

- (72) *Carpenter Simeon Alexis, who suffered a chest wound when the crane collapsed, is reported to be awake and speaking here at New York Hospital and is expected to recover. His colleagues are maintaining vigil, among them foreman Scott Bair. He says,*
[he himself escaped the crane collapse]
barely and is physically okay but feels like a mess.

(72') [[[[he]_{F1} ~1] [himself]_{F2} [escaped]_{F2} the crane collapse ~ 2]
k = 'A worker of Bair's was injured in the crane collapse'
f = $\exists u_{ee} \exists P_{et-} [u(\text{Blair}) P\text{'ed the crane collapse}]$

The *f0* peak contour is consistent with a double focus configuration. The two focus associates, *himself* and *escaped* are realized with the highest *f0* peaks in the domain corresponding to the scope of focus. That the *f0* peak realized on *escaped* is higher than the *f0* peak on *himself* is not expected from Eady *et al.*'s experimental results, but the lack of downstepping is sufficient to distinguish the *f0* contour as belonging to a double focus configuration rather than an early single focus configuration or a clausal focus configuration. The *f0* contour is not consistent with a first focus associate on the subject since the *f0* peak on *he* is lower than the *f0* peak on *himself*.

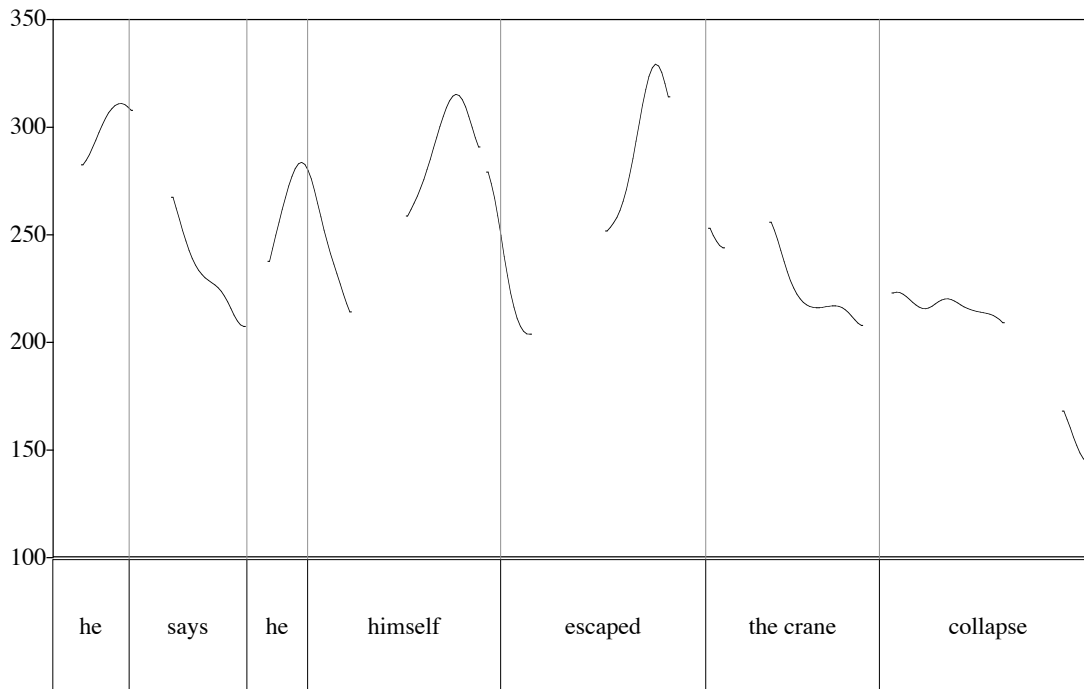


Figure 16. Smoothed *f0* track of (52). Double focus configuration (adnominal ER, predicate).

In (73), NPR radio host Robin Young interviews Stephen Heffner about the online political convention he created, *Jesus in 2008*, which asked delegates what Jesus would do if he were a presidential candidate in the U.S. in 2008. At the start of the interview, Young introduces the topic with a recording of real presidential candidate Mike Huckabee being asked what position Jesus would take on the death penalty. In the discussion that follows, Heffner and Young each use the predicates *in favor (of) the death penalty* and *against the death penalty* applied to the individual *Jesus*. There is thus a very salient discourse antecedent of the form ‘x is *P* of the death penalty’ where x is an individual and *P* is predicative adjective. This antecedent would be sufficient to license a double focus configuration in which the initial focus was the subject argument (i.e. *he himself*).

In suggesting the political danger of taking a position on the death penalty, Young draws on our world knowledge that Huckabee is a Republican candidate, and that his Republican supporters are likely to favor the death penalty. She introduces a salient discourse antecedent *Supporters of Huckabee are in favor of the death penalty*, which contains an alternative relational predicate-adjectival predicate pair: <SUPPORTERS-OF(), in-favor-of()> .

- (73) Heffner: *We’re trying to build a platform here with Jesus as the standard, against which we measure the decisions and proposals we make on all these issues [...]*
- Young: *But I’m wondering, are you finding that people see Jesus differently? For instance, we’ve had Mary Luti on the program [...] In her sermon the next Sunday she said she thought Jesus’s position on the death penalty would be clear. He would say that he’d suffered it, as did anyone else on Calvary Hill; **he’d be against it**. [...] Have you found people who say **Jesus would be in favor of the death penalty**?*
- Heffner: *We might have had one delegate weigh in so far, saying that **he might favour it**. But we had someone come in and say exactly what you just said which is that **he, more than anyone, would be against it**, having suffered it himself.*

Young: *We began this discussion, again, by citing Mike Huckabee, who didn't directly answer the question that was asked him ["What would Jesus do about the death penalty?"], but maybe that's because it was politically dangerous for him. As it turns out,*
[he himself is very much against the death penalty]

- (73') [[[[he]_{F1} ~1] [himself]_{F2} is very much [against]_{F2} the death penalty] ~ 2]
k = 'Huckabee's supporters are in favor of the death penalty'
f = $\exists u_{ee} \exists P_{eet} [u(\text{Huckabee}) \text{ is } P \text{ the death penalty}]$

The *f0* contour for (73) is consistent with a double focus configuration. The two focus associates, *himself* and *against* are realized with the highest *f0* peaks in the domain corresponding to the scope of focus. The *f0* contour is inconsistent with the clausal focus configuration since the lexical words *death* and *penalty* have reduced *f0*. The *f0* is inconsistent with a configuration of double focus and a subject first focus associate because *himself* is realized with a higher *f0* peak than *he*.

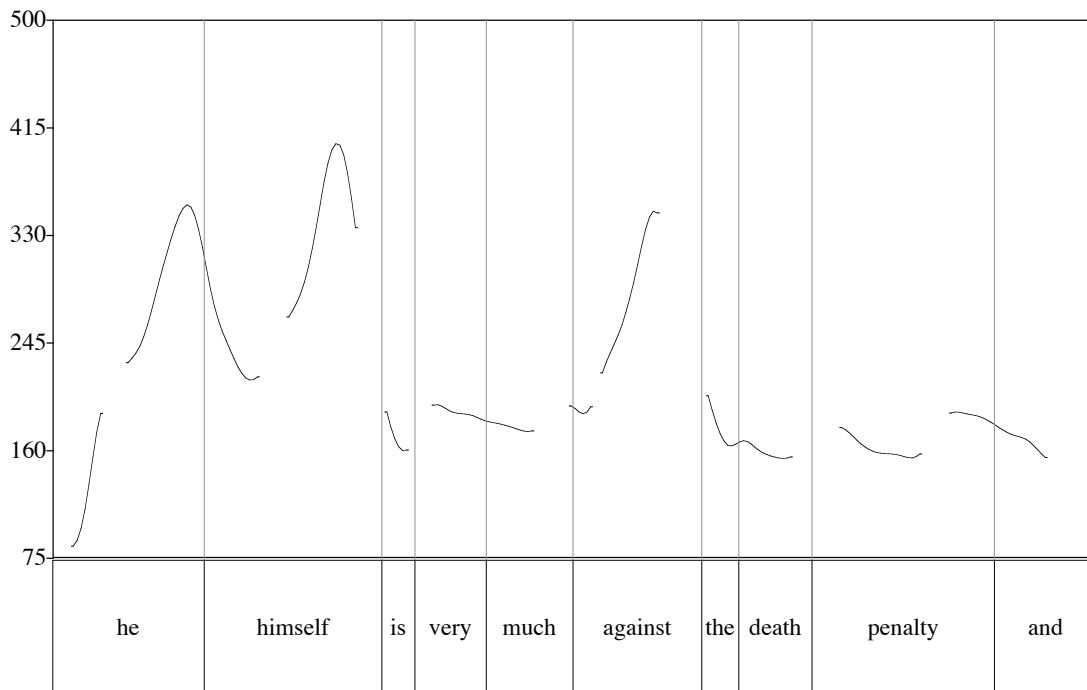


Figure 17. Smoothed *f0* track (in Hz) of (52). Double focus configuration (adnominal ER, predicate).

In (57), a newscaster explains that the Philipp Lahm Foundation is having running shoes sent to Ethiopia. As Creswell observed in her written corpus, many examples of the double focus configuration are licensed in part by an alternative predicate with opposite polarity. In this example, we understand that there are people working on behalf of or in collaboration with Philipp Lahm who have, in fact, been to Ethiopia. Thus, *has been to Ethiopia* is a salient alternative to the negative predicate *has not been to Ethiopia* and *people working on behalf of* is a salient alternative relational predicate. The target sentence is licensed by the alternative relational predicate-verbal predicate pair: <REPRESENTATIVES-OF(), be-in-Ethiopia(>

- (74) *For the Bokoji project, the Philipp Lahm Foundation is providing the running shoes as well as organizing the transport to Ethiopia. The football star says that the project fits perfectly into the portfolio of his foundation*
 [and that while **he himself** has so far not been to Ethiopia]
he is eager to maybe go there in the course of the project

- (74') [[[[he]_{F1} ~ 1] [himself]_{F2} has so far [not]_{F2} been to Ethiopia] ~ 2]
 $k =$ 'People working on behalf of Philipp Lahm have been to Ethiopia'
 $f = \exists u_{\langle ee \rangle} \exists P_{\langle et \rangle} . [u(\text{Philipp Lahm}) \text{ has } P'ed]$

The f_0 peak contour of the utterance is almost exactly as predicted for a double focus configuration, with nearly equal height on the two highest peaks. The associate *he* has an f_0 peak, but its height is lower than the f_0 on the adnominal ER, consistent with narrow focus on the adnominal ER. *Stress-F* is satisfied, since the two associates are realized with greatest intonational prominence in the domain corresponding to the scope of focus.

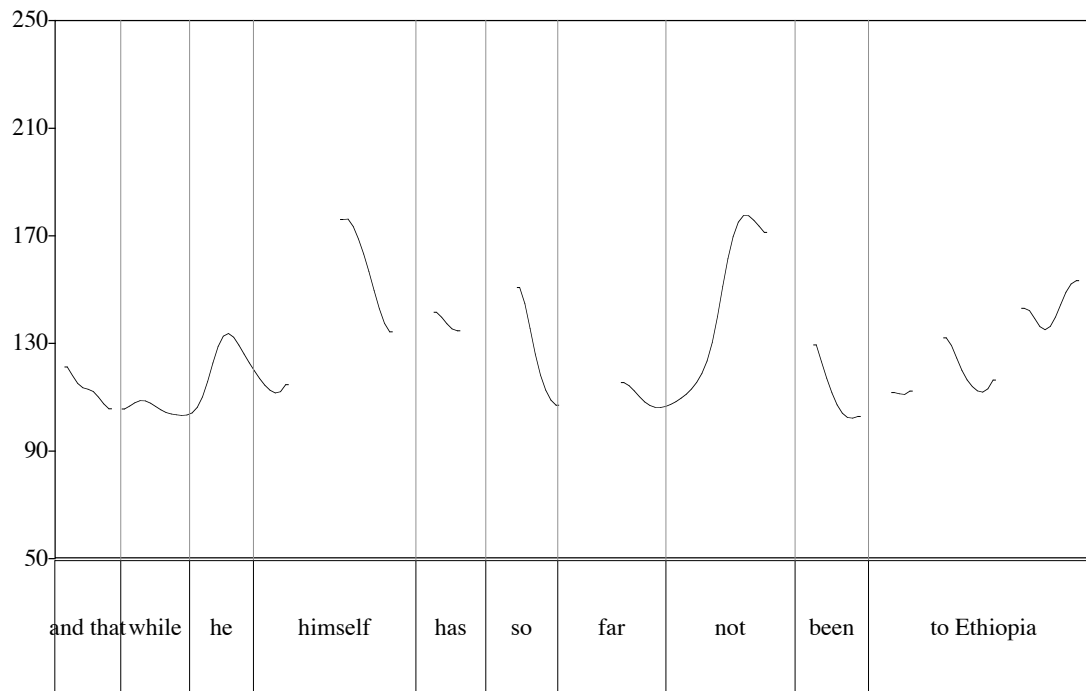


Figure 18. Smoothed f_0 track of (74). Double focus configuration (adnominal ER, predicate).

Finally, I end this section with examples of the double focus configuration with initial focus on the subject. In (75), Gene Epstein, economics editor at *Barons*, expounds on the possible dangers of galloping inflation. Certain federal reserve chairmen believe that galloping inflation is an economic situation that will return if they don't intervene. These chairmen are salient alternatives to *Greenspan* and their attitude of indifference to the proposition *galloping inflation will return* is a salient alternative to the attitude predicate *fear*. Double focus on the target sentence is licensed by a salient antecedent *Some federal reserve chairmen don't care whether galloping inflation will return*, which contains the alternative individual-predicate pair <some federal reserve chairmen, don't care()>.

- (75) *And thirdly, here, this has been an issue we haven't explored, we somewhat differ over what's happened in the last 20 years. I absolutely agree with you, by the way, that Paul Volker started the high priesthood of federal reserve chairman who care about making sure that galloping inflation doesn't return. But Greenspan, and now Bernanke have greatly been helped by the end of the cold war, which caused a massive disinflation around the world by the increase in productivity, which also helped cause disinflation. And then Greenspan, in the only really valuable part of his recent book, points out that by 2030, the future federal reserve chairman is gonna feel a lot of pressure to expand the money supply and*
[he himself fears that galloping inflation will return]
so that's a third reason to be concerned

- (75') [[[[[he]_{F1} ~1] himself]_{F2} [fears]_{F2} that galloping inflation will return] ~ 2]
 $k =$ 'Some federal reserve chairmen don't care whether galloping inflation will return'
 $f = \exists x_e \exists P_{\langle \text{ett} \rangle} [x \text{ } P \text{'s that galloping inflation will return}]$

The $f0$ contour is consistent with a double focus configuration: the two associates *he himself* and *fears* are realized with the highest $f0$ peaks within the domain corresponding to the scope of focus. Within the subject *he himself*, both *he* and *himself* are realized with $f0$ peaks; the peak on *himself* is slightly lower, as predicted. The $f0$ contour following the second associate *fears* is reduced since the pitch accent on *fears* is nuclear. This $f0$ contour is not consistent with an adnominal ER first focus associate; this would require *himself* to be realized with a higher $f0$ peak than *he*. Note that there are a couple artifacts at the end of the words *inflation* and *return*; these are not true $f0$ peaks.

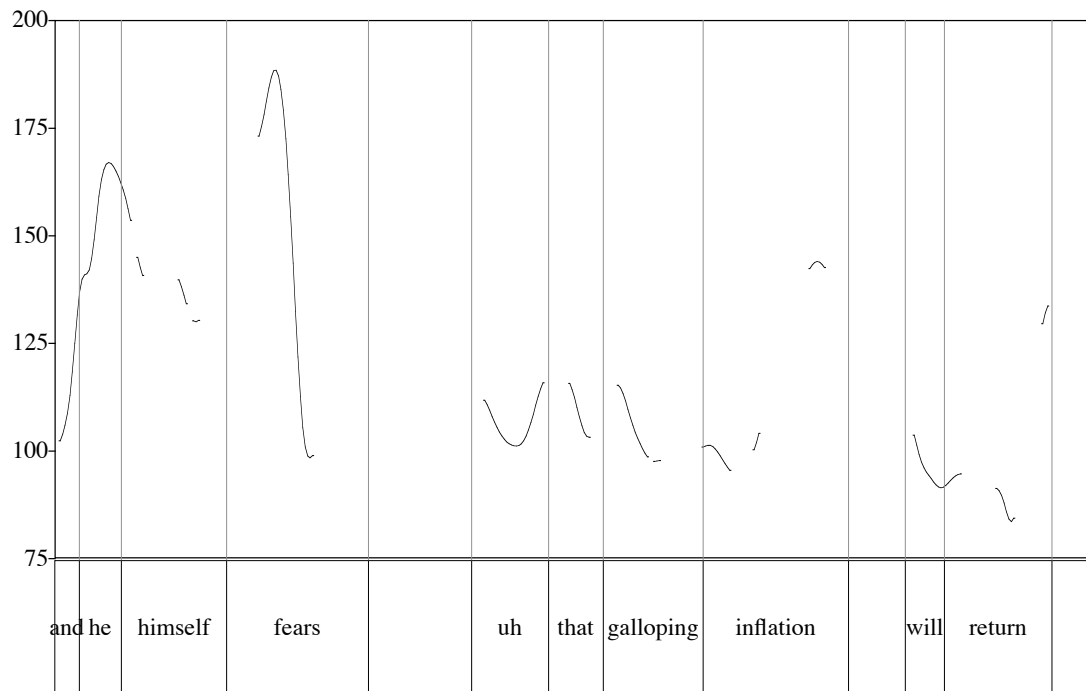


Figure 19. Smoothed f_0 track (in Hz) of (75). Double focus configuration (subject, predicate).

In (76), political author and editor Linda Bridges discusses the failed presidential campaign of U.S. Senator John McCain. Bridges suggests that McCain's campaign was not inspiring, while the interviewer maintains that McCain is an inspiring figure. By introducing the negated predicate *not be an inspiring campaign*, Bridges makes salient the positive predicate *be an inspiring campaign*, which is contrasted with *inspiring figure*, and McCain is likewise contrasted with his campaign. A double focus configuration for the target sentence is licensed by a discourse antecedent *The campaign was an inspiring campaign*, which contains the individual-predicate pair <the campaign, be a campaign(>).

- (76) Bridges: *We wind up with John McCain, who wound up bouncing from one thing to another, his campaign was in disarray. And then of course the coup de grace was the financial meltdown. But even before that it wasn't a very inspiring campaign.*
- Interviewer: *To put it mildly. Even though it had the potential to be. Because [he himself is an inspiring figure]*

(76') [[[[[he]_{F1} ~1] himself]_{F2} is an inspiring [figure]_{F2}] ~ 2]
 k = 'his campaign was an inspiring campaign'
 f = $\exists x_e \exists P_{et}$ 'x was an inspiring P'

The discourse could also license double focus with an adnominal ER first focus, since the antecedent *The campaign of John McCain was an inspiring campaign*, with the alternative relational predicate-verbal predicate pair <CAMPAIGN-OF(), be a campaign()>, is also salient. This configuration, however, is not supported by the speaker's f_0 contour: within the subject *he himself*, *he* has a higher f_0 peak. The two focus associates, *he himself* and *figure* are realized with the highest f_0 peaks in the domain corresponding to the scope of focus.

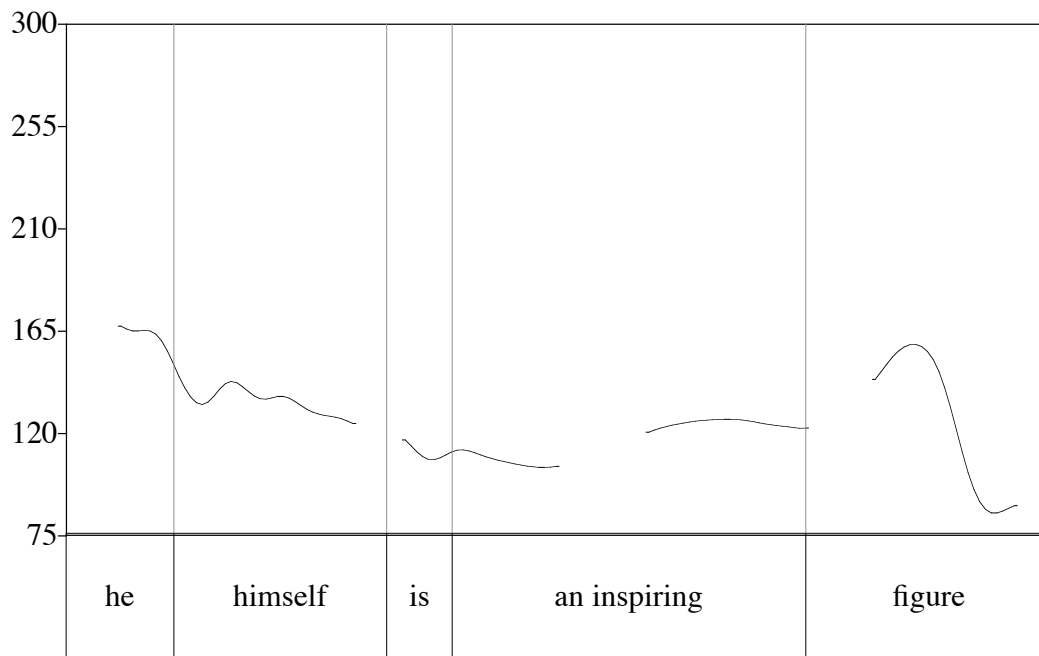


Figure 20. Smoothed f_0 track (in Hz) of (76). Double focus configuration (subject, predicate).

In summary, we see that the distribution of examples with double focus mirrors the configurations of narrow focus on the subject *he himself* and narrow focus on the adnominal ER *himself*. The examples of double focus with subject first focus are predicted by the extended

FOCUS SENSITIVE OPERATOR approach, but not by the FOCUSED ASSERTION OF IDENTITY approach. The examples of double focus with adnominal ER first focus are predicted by both approaches.

The second focus associate in the double focus associate may be the entire predicate or some sub-constituent of the predicate. The results of Eady *et al.* predict the *f0* peak on the second associate should be lower than the *f0* peak on the first associate. In the corpus examples, the second associate was realized with an *f0* peak that was slightly higher. The important points are (1) that the two associates are realized with the highest *f0* peaks within the domain corresponding to the scope of focus and (2) that the second focus associate is realized with an *f0* peak that is not downstepped: i.e. that the *f0* peak is greater than it would otherwise be in an clausal focus configuration.

7 *Contrastive topic*

In order to represent the contrast between single late focus and double focus, I adopted the ToBI convention of boundary tones to indicate prosodic phrasing. Until now, I have also assumed that pitch accents come in only one type, a simple high pitch accent H*. In this section, I consider the possible contribution of other accents.

One kind of contour in particular—often referred to as a “B-accent” (cf. Jackendoff 1972)⁴ or “contrastive-topic accent” (cf. Büring 2003)—is claimed to carry an additional pragmatic implicature, such as lack of speaker commitment (Hirschberg & Ward 1985), set membership (Ladd 1980) or an open, disputable question (Büring 2003). In (79-82), for example, we have the intuition that B does not completely answer A’s question.

- (77) A: Did you feed the animals? (Ladd 1980)
 L*(+H) L-H%
 B: I fed [the cat]_F

⁴ The A-accent / B-accent terminology is originally due to Bolinger. Ladd (1980:216) points out that in Bolinger’s system, fall-rise is in fact labeled an A-accent.

- (78) A: Do you want a glass of water? (Ladd 1980)
 L*(+H) L-H%
 B: I'll have [a beer]_F
- (79) A: How did the exam go? (Rooth 1992)
 L*(+H) L-H%
 B: Well, [I]_F passed
- (80) A: Will Uncle Michael and Aunt Carolyn be coming to the rehearsal dinner? (Rooth 1992)
 L*(+H) L-H%
 B: They're invited

I will follow Ward & Hirschberg in distinguishing more than one kind of fall-rise contour. In ToBI conventions (Beckman & Ayers 1994), the fall-rise contours in (77-80) are realized with a low nuclear accent and a high boundary tone: L*(+H)L-H%. Pragmatically, this fall-rise contour carries the described implicature. I do not have anything to say about which of the proposed pragmatic accounts is correct; the important point here is that this particular contour comes with such an implicature.

The other fall-rise contour is the result of a low phrase boundary tone followed by a high intermediate phrase boundary: -L H%. In the double focus constructions described in Section 7, the two focus associates are realized with nuclear pitch accents. In fast speech, it is normal to produce the two pitch accents as the last within intermediate phrases, requiring only a intermediate phrase boundary tone (e.g. -L in 81a). In more deliberate or emphatic speech, one might also realize the nuclear pitch accents within separate intonational phrases, requiring an intonational phrase boundary tone (e.g. H% in 81b; Beckman 1996). It is the separate intonational phrases that result in a fall-rise contour occurring on the first nuclear accent. The fall-rise contour in (81b) contour does not introduce an implicature.

- (81) ‘Who heard what?’
- | | | | |
|----|----------------------|--------------------------------|---------|
| | H*-L | | H* L-L% |
| a. | [[[Ben] _F | heard [shots] _F] ~ | |
| | H*-L H% | | H* L-L% |
| b. | [[[Ben] _F | heard [shots] _F] ~ | |

The fall-rise contour without implicature is not specific to the double focus configuration; as noted by Wagner (2006), it may in fact occur with any syntax-induced prosodic boundary, such as co-ordination. In (82), a utterance-level boundary tone is used to indicate co-ordination between the first argument *Mary* and the second and third arguments *Sue and Jane*.⁵

- (82) A: Who came to the party?
- | | | | |
|----|----------|------------|---------------|
| | H*-L H% | | H*L-L% |
| a. | B: Mary, | or | Sue and Jane. |
| b. | B: Mary, | and | Sue and Jane. |

Eckardt (2001) claims that all examples of the “no-surprise” uses of the adnominal (cf. “big shot” uses) are realized with a rise-fall contour on the adnominal ER and a rise on a second focus associate.⁶ Her example (83) is introduced with a pair-list context. I believe that the target sentence in this context is most likely to be realized with a double focus configuration, since there are two salient antecedents: *the archbishop of the king wore a mitre* which contains the relational predicate-argument pair <ARCHBISHOP-OF(), mitre> and *the lords of the king wore shining helmets*, which contains the relational predicate-argument pair <LORDS-OF(), shining helmets>. The fall-rise contour without implicature may occur with this focus configuration if it is produced slowly or emphatically such that there are two intonational phrases. The other fall-

⁵ The ToBI transcription is mine; Wagner (2010) labels the intonation on *Mary* as rise-fall-rise.

⁶ As discussed in Section 2, this sentence can also be realized with “all new” prosody in a clausal focus configuration.

(83') *The Royal Parade was held today and, as expected...*

H*	H*	H*	H*	L-L%	“all new” prosody
[The king himself	wore a crown] _F ~				

rise contour may also be possible, if the context is such that the speaker's utterance is not a complete answer. For instance, suppose we know that the jester was also present and that he wears flamboyant headgear. If I intend to continue my list of courtier-headgear pairs, I might signal that the list is incomplete with the L* L-H% contour.

- (83) *The archbishop was easy to spot, thanks to his mitre. His Lords wore shining helmets...*

	L*(+H)L-H%	H* L-L%	contrastive topic prosody
	H* L-H%	H* L-L%	double focus prosody
[The king [himself] _{F2}	wore [a crown] _{F2}]	~2

In (83), the fall-rise contour is realized on the adnominal ER. The fall-rise contour may also be realized over the subject, as in (84). In order create examples that disfavor narrow focus on the adnominal ER, I again choose individuals who are unlikely to stand in a salient relation to each other: the President, the Pope and Lady Gaga.

- (84) a. The President wore Armani, the Pope wore Calvin Klein and...

L*	L*	L*(+H) L-H%	H* L-L%
H*	H*	H* L-H%	H* L-L%
[[Lady Gaga herself] _{F2}		wore [Versace] _{F2}]

~2

- b. The Pope wore Calvin Klein, Lady Gaga wore Versace and...

L*	L*(+H) L-H%	H* L-L%
H*	H* L-H%	H* L-L%
[[The President himself] _{F2}	wore [Armani] _{F2}]

~2

- c. The President wore Armani, Lady Gaga wore Versace and...

L*	L*(+H) L-H%	H*	H* L-L%
H*	H* L-H%	H*	H* L-L%
[[The Pope himself] _{F2}	wore [Calvin Klein] _{F2}]	~2

Just as the mere presence of an adnominal ER in an utterance does not determine the focus configuration of the utterance, so too the presence of an adnominal ER does not entail a

particular pitch accent. As we can observe in (85), the same pitch contour is available in utterances lacking the adnominal ER.

- (85) a. The President wore Armani, the Pope wore Calvin Klein and...

L*	L*(+H)	L-H%	H*	L-L%
H*	H*	L-H%	H*	L-L%
[[Lady Gaga] _{F2}			wore [Versace] _{F2}] ~2	

- b. The Pope wore Calvin Klein, Lady Gaga wore Versace and...

L*(+H)	L-H%	H*	L-L%
H*	L-H%	H*	L-L%
[[The President] _{F2}		wore [Armani] _{F2}] ~2	

- c. The President wore Armani, Lady Gaga wore Versace and...

L*(+H)	L-H%	H*	H*	L-L%
H*	L-H%	H*	H*	L-L%
[[The Pope] _{F2}		wore [Calvin Klein] _{F2}] ~2		

Unfortunately, none of the adnominal ER tokens in the web corpus appear to be realized with fall-rise contour, as identified either by prosodic or pragmatic criteria. Naturally, it does not follow from the absence of this contour in the corpus that the adnominal ER is never realized in this way. It may be the case such examples have simply been misclassified, or that the contour simply has a low frequency. Since the focus semantics of the adnominal ER and the focus semantics of the sentence that contains it are independent on the extended FOCUS SENSITIVE OPERATOR analysis, nothing should prevent a realization of the adnominal ER with either of the fall-rise contours. A more rigorous investigation of fall-rise contours with adnominal ERs requires laboratory investigation. I leave this for future research.

8 *Predicate focus and overlapping associate focus*

Two logically possible focus configurations remain: predicate focus (or late single focus in the terminology of Eady *et al.*) and associate focus. In the predicate focus configuration, neither

the adnominal ER, nor the argument to which the adnominal ER belongs, is the associate of a clause-level focus; rather the predicate or some sub-constituent of the predicate is the associate of a clause-level focus. In (86), *chair* is the focus associate. The focus configuration in (86b) is licensed by the salient discourse antecedent ‘he himself will be a member of the committee’.

- (86) a. I heard the Provost himself will be a member of the committee?
 b. [[[[he]_{F1} ~1] himself] will [**chair**]_{F2} the committee] ~ 2

- (86') $k =$ ‘The Provost himself will be a member of the committee’
 $f = \exists P_{\langle \text{eet} \rangle}.$ [the Provost will P the committee]

The focus associate *chair* will be realized with the highest $f\theta$ peak in the domain corresponding to the domain of focus, in order to satisfy *Stress-F*; the $f\theta$ contour in the rest of the utterance will be reduced.

- (87) [[[[he]_{F1} ~1] himself] will [**chair**]_{F2} the committee] ~ 2

An attested occurrence of this pattern from the corpus is transcribed in (88). The speaker is discussing the charitable foundation of actor Michael J. Fox. The focus associate of clause-level focus is the verb *contracted*. The focus configuration on the target utterance is licensed by the salient discourse antecedent ‘Michael J. Fox set up a foundation for Parkinson’s disease’.

- (88) *You’ve probably heard of the Michael J. Fox foundation, set up by the actor Michael J. Fox*
 [for Parkinson’s disease because **he himself** contracted the disease]

- (88') [[[[he]_{F1} ~1] himself] [contracted]_{F2} the disease] ~ 2
 $k =$ ‘Michael J. Fox himself set up a foundation for Parkinson’s disease’
 $f = \exists P$ ‘Michael J. Fox himself P’ed Parkinson’s disease’

The focus associate *contracted* is realized with the highest $f\theta$ within the domain corresponding to the scope of focus, the clause *he himself contracted the disease*. The rest of the

material is realized with reduced f_0 contour. The reduced f_0 on *himself* is clearly inconsistent with a configuration narrow focus on the adnominal ER, as predicted on the FOCUS SENSITIVE OPERATOR approach.

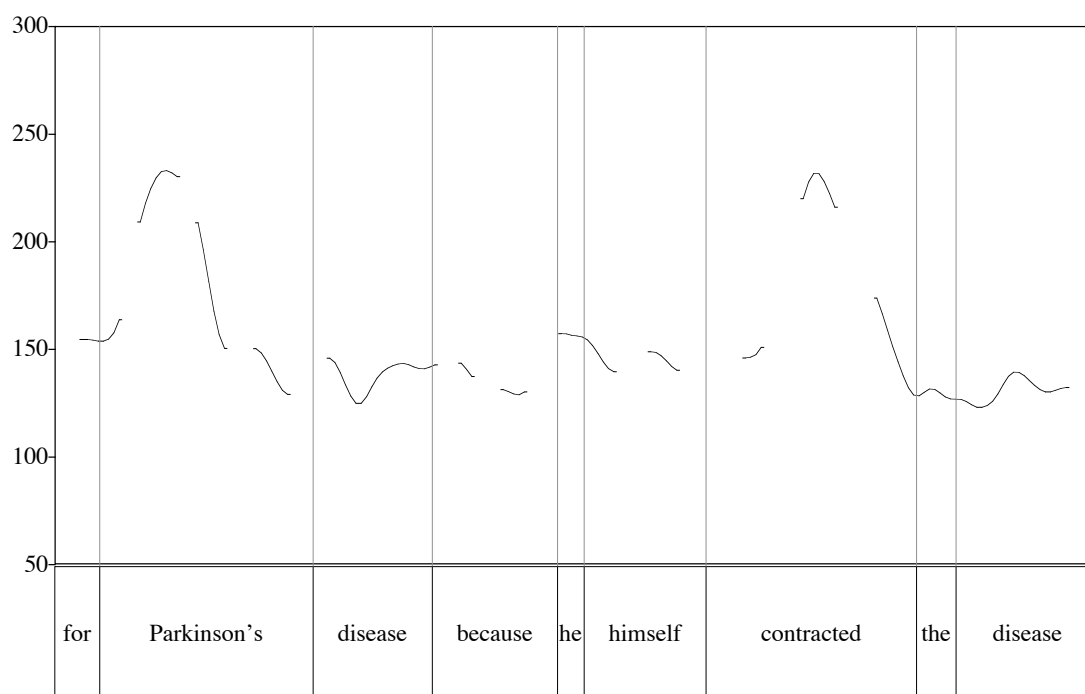


Figure 21. Smoothed f_0 track of (88). Predicate focus configuration.

Lastly, it should also be possible for a clause-level focus operator to associate with the associate of the adnominal ER. For example, *he* in (89) is the focus associate of both the focus sensitive operator *himself* and a clause-level focus operator. Accordingly, I call this configuration overlapping associate focus. The associate *he* must be most prominent within both the domain corresponding to the associate and the domain corresponding to the clause. *He* will have the highest f_0 peak, and the following material will be realized with a reduced f_0 contour. The clause-level focus in this configuration is highly metalinguistic, being appropriate only as a correction (cf. 89) or in response to echo questions (cf. 90). This configuration is predicted by

the extended FOCUS SENSITIVE OPERATOR approach, but not the FOCUSED ASSERTION OF IDENTITY approach.

I find no examples of this configuration in the web-harvested corpus. Of course, it does not follow from the absence of this configuration that it does not occur at all; more likely a very low frequency of occurrence.

- (89) a. The mayor himself will chair the committee.
 b. No, [[[**he**]_{F1,F2} ~1] himself] will chair the committee] ~ 2
 [uttered while pointing at the Provost]

- (90) a. Who himself will chair the committee?
 b. [[[[**he**]_{F1,F2} ~1] himself] will chair the committee] ~ 2

— — — — —

- (91) [[[[**he**]_{F1,F2} ~1] himself] will chair the committee] ~ 2

9 Conclusion

In the introduction, I explained that the formal semantics literature on adnominal ERs is divided into two main camps: those who believe that the adnominal ER is deterministically focused (the FOCUSED ASSERTION OF IDENTITY approach and to a lesser extent the DOUBLE OPPOSITION approach) and those who believe that the adnominal ER is in fact a focus operator (the FOCUS SENSITIVE OPERATOR approach). The former approaches predict that only one focus configuration is possible for utterances containing the adnominal ER—two configurations if we allow both single and double focus configurations with narrow focus on the adnominal ER.

Having hypothesized 9 different configurations (and possibly more if we allow for fall-rise contours) predicted by the non-deterministic approach, I identified naturally-occurring examples of at least 7. We discovered that it possible to distinguish the prosody of these configurations using this minimal representation of prominence as *f0* peak contours. And we can

model the semantics of these configurations using discourse antecedence and entailment relations

The 9 different configurations are summarized in (92), with focus structure, antecedence requirement and a representation of *fθ* topline.

(92) a. Clause focus

—
—
—
—

[[[[he]_{F1} ~1] himself] will chair the committee]_{F2} ~ 2
 Clause-level antecedent: Any proposition of type <*t*>

b. Argument focus (subject)

—
—
—
—

[[[[he]_{F1} ~1] himself]_{F2} will chair the committee] ~ 2
 Clause-level antecedent: Proposition of the form ‘*x* will chair the committee’

c. Adnominal ER focus

—
—
—
—

[[[[he]_{F1} ~1] [himself]_{F2}] will chair the committee] ~ 2
 Clause-level antecedent: Proposition of the form ‘*u*(he) will chair the committee’

d. Double focus (Subject, predicate)

—
—
—
—

[[[[he]_{F1} ~1] himself]_{F2} will [chair]_{F2} the committee] ~ 2
 Clause-level antecedent: Proposition of the form ‘*x* will *P* the committee’

- e. Double focus (Adnominal ER, predicate)

—
—
—

[[[[he]_{F1} ~1] [**himself**]_{F2}] will [**chair**]_{F2} the committee] ~ 2
 Clause-level antecedent: Proposition of the form ‘*u*(he) will *P* the committee’

- f. Sub-clausal focus (argument)

—
—

[[[[**he**]_{F1} ~1] **himself**]_{F2} ~2] and [[his committee]_{F3} ~3]
 Constituent antecedent: (Set of) individual(s) of form ‘*x*’

- g. Sub-clausal focus (adnominal ER)

—
—

[[[he]_{F1} ~1] [**himself**]_{F2} ~2] and [his [committee]_{F3} ~3]
 Constituent antecedent: (Set of) individual(s) of form ‘*u*(he)’

- h. Predicate focus

—
—
—

[[[[he]_{F1} ~1] himself] will [**chair**]_{F2} the committee] ~ 2
 Clause-level antecedent: Proposition of the form ‘he himself will *P* the committee’

- i. Overlapping associate focus

—
—
—

[[[[**he**]_{F1,F2} ~1] himself] will chair the committee] ~ 2
 Clause-level antecedent: Proposition of the form ‘*x* himself will chair the committee’

Certainly, many empirical questions remain, and I noted on several occasions questions which require experimental investigation of data elicited in the laboratory. For example, I have knowingly ignored the existence of pitch accents other than H*. Additional configurations which use more than just the H* accent, such as contrastive topic discussed in Section 7, may well occur. In order to make the necessary distinctions, these may require richer phonological representations such as ToBI, and/or gradient representations which transcribe prominence without reference to pitch accent type, such as RPT (Mo 2010).

In some examples, such as double focus, the relation of *f0* peak heights were sufficient to distinguish between configurations, but they deviated from the relation predicted. Is this variation grammatically significant? One can manipulate *f0* with synthesis to compare this variation in minimal n-tuples and elicit judgments in laboratory perception experiments.

I noted in many examples that the discourse could support more than one focus configuration for an utterance. To what extent may a focus configuration be optional in a given context? One could ask participants to read written transcripts, or constructed examples, with context and compare prosodic realizations. Similarly, one could manipulate the *f0* of a target utterance and elicit judgments of discourse congruence in laboratory perception experiments.

For some configurations, such as predicate focus, I identified very few examples. I do not know why some configurations occur more frequently than others. Are some focus configurations easier to accommodate than others? One could ask participants to read written transcripts, or constructed examples, without context and compare prosodic realizations.

The principle insight of this chapter, however, did not require any such laboratory investigation. Indeed, it is work relying on introspective and experimental data that overlooked this insight: that the adnominal ER may occur within several distinct focus configurations, each with a distinct prosodic realization. The prediction of focal determinism—that only one focus configuration is possible (cf. 1)—is simply untenable.

Focus and prominence on the adnominal ER are not deterministic; rather, they vary according to context. Consequently, there is no evidence that the adnominal ER is lexically specified for focus (inherent focus) or that it is semantically and pragmatically vacuous and only meaningful when focused (obligatory focus). From the perspective of theory, this is happy result. Stipulating focus effects for individual constructions is not maximally explanatory, and can support at best an intermediate, and possibly a weak theory of semantic interpretation, in the sense of Rooth (1992).

From a practical standpoint, the burden of proof for other constructions lies, I believe, with those who allege they are deterministically focused. They must demonstrate that the normal variation of association with focus for utterances containing the construction fails to obtain.

Moreover, theorists and experimentalists alike stand to gain considerable insights from careful examination of naturally occurring speech data, even if only to provide a reality check for one's intuitions or inform experimental design.

APPENDIX A

ACOUSTIC MEASURES

For each utterance of “than I did”, the following phonetic segments were annotated: V1, the vowel [æ] of *than*; N1, the nasal [n] of *than*; V2, the diphthong [aɪ] of *I*; C3, the stop closure and burst of the initial [d] in *did*; and V3, the vowel [ɪ] of *did*.

Acoustic measure	Description
duration [V1,V2,V3,C3]	duration of segment (vowel, stop closure)
loudness/intensity	
meanIntensity_[V1,V2,V3]	RMS Intensity over vowel
meanIntensity_ratio	meanIntensity_V2 to meanIntensity_V3
maxIntensity_[V1,V2,V3]	max RMS Intensity
maxIntensity_ratio	maxIntensity_V2 to maxIntensity_V3
minIntensity_[V1,V2,V3]	min RMS Intensity
minIntensity_ratio	minIntensity_V2 to minIntensity_V3
rangeIntensity_[V1,V2,V3]	range of RMS Intensity in vowel
rangeIntensity_ratio	rangeIntensity_V2 to rangeIntensity_V3
maxIntTime_[V1,V2,V3]	time of intensity max relative to vowel duration
maxIntTime_ratio	ratio of maxIntTime_V2 to maxIntTime_V3
minIntTime_[V1,V2,V3]	time of intensity min relative to vowel duration
minIntTime_ratio	ratio of minIntTime_V2 to minIntTime_V3
energy_[V1,V2,V3]	mean energy over vowel
energy_ratio	ratio of mean energy of V2 to mean energy of V3
power_[V1,V2,V3]	mean power of vowel
power_ratio	ratio of mean power of V2 to mean power of V3
amp_[V1,V2,V3]	mean amplitude of vowel
amp_ratio	ratio of mean amplitude of V2 to mean amplitude of V3
voice	
pulses_[V1,V2,V3]	number of glottal pulses
jitter_[V1,V2,V3]	jitter
shimmer_[V1,V2,V3]	shimmer
pulses_ratio	ratio of V2 pulses to V3 pulses
jitter_ratio	ratio of V2 jitter to V3 jitter
shimmer_ratio	ratio of V2 shimmer to V3 shimmer
fundamental frequency (<i>f</i> 0)	
<i>f</i> 0 [V1,V2,V3]	mean <i>f</i> 0 of vowel
mean <i>f</i> 0_ratio	ratio of mean <i>f</i> 0 of V2 to mean <i>f</i> 0 of V3
max <i>f</i> 0 [V1,V2,V3]	max <i>f</i> 0 of vowel
max <i>f</i> 0_ratio	ratio of max <i>f</i> 0 of V2 to max <i>f</i> 0 of V3
min <i>f</i> 0 [V1,V2,V3]	min <i>f</i> 0 of vowel
min <i>f</i> 0_ratio	ratio of min <i>f</i> 0 of V2 to min <i>f</i> 0 of V3
max <i>f</i> 0Time [V1,V2,V3]	time of <i>f</i> 0 max relative to vowel duration
max <i>f</i> 0Time_ratio	ratio of max <i>f</i> 0Time of V2 to V3
min <i>f</i> 0Time [V1,V2,V3]	time of <i>f</i> 0 min relative to vowel duration
min <i>f</i> 0Time_ratio	ratio of min <i>f</i> 0Time of V2 to V3
range <i>f</i> 0 [V1,V2,V3]	<i>f</i> 0max - <i>f</i> 0min

range <i>f0</i> ratio	ratio <i>f0</i> max V2 to <i>f0</i> max V3
first formant (f1)	
maxf1_[V1,V2,V3]	max f1 of vowel
maxf1_ratio	ratio of max f1 of V2 to max f1 of V3
minf1_[V1,V2,V3]	min f1 of vowel
minf1_ratio	ratio of min f1 of V2 to min f1 of V3
maxf1Time_[V1,V2,V3]	time of f1 max relative to vowel duration
maxf1Time_ratio	ratio of maxf1Time of V2 to V3
minf1Time_[V1,V2,V3]	time of f1 min relative to vowel duration
minf1Time_ratio	ratio of minf1Time of V2 to V3
rangef1_[V1,V2,V3]	f1max - f1min
rangef1_ratio	ratio f1max_V2 to f1max_V3
f1TimeIntmax_[V1,V2,V3]	f1 value at time of intensity max
f1Time <i>f0</i> max_[V1,V2,V3]	f1 value at time of <i>f0</i> max
f1Time[10,20... 90]_[V1,V2,V3]	f1 value at 10% 20%... 90% of vowel duration
f1bandIntmax_[V1,V2,V3]	f1 bandwidth value at time of intensity max
f1band <i>f0</i> max_[V1,V2,V3]	f1 bandwidth value at time of <i>f0</i> max
f1band[10,20... 90]_[V1,V2,V3]	f1 bandwidth value at 10% 20%... 90% of vowel duration
second formant (f2)	
maxf2_[V1,V2,V3]	max f2 of vowel
maxf2_ratio	ratio of max f2 of V2 to max f2 of V3
minf2_[V1,V2,V3]	min f2 of vowel
minf2_ratio	ratio of min f2 of V2 to min f2 of V3
maxf2Time_[V1,V2,V3]	time of f2 max relative to vowel duration
maxf2Time_ratio	ratio of maxf2Time of V2 to V3
minf2Time_[V1,V2,V3]	time of f2 min relative to vowel duration
minf2Time_ratio	ratio of minf2Time of V2 to V3
rangef2_[V1,V2,V3]	f2max - f2min
rangef2_ratio	ratio f2max_V2 to f2max_V3
f2TimeIntmax_[V1,V2,V3]	f2 value at time of intensity max
f2Time <i>f0</i> max_[V1,V2,V3]	f2 value at time of <i>f0</i> max
f2Time[10,20... 90]_[V1,V2,V3]	f2 value at 10% 20%... 90% of vowel duration
f2bandIntmax_[V1,V2,V3]	f2 bandwidth value at time of intensity max
f2band <i>f0</i> max_[V1,V2,V3]	f2 bandwidth value at time of <i>f0</i> max
f2band[10,20... 90]_[V1,V2,V3]	f2 bandwidth value at 10% 20%... 90% of vowel duration
ratia	
f1f2TimeIntmax [V1,V2,V3]	f2-f1 at time of intensity max
f1f2Time <i>f0</i> max [V1,V2,V3]	f2-f1 at time of <i>f0</i> max
f1f2Time[10,20... 90] [V1,V2,V3]	f2-f1 value at 10% 20%... 90% of vowel duration
spectral measures	
h1minush2p0_[V1,V2,V3]	1st harmonic minus 2nd harmonic at time of <i>f0</i> maximum
h1minush3p0_[V1,V2,V3]	1st harmonic minus 3rd harmonic at time of <i>f0</i> maximum
h2minush3p0_[V1,V2,V3]	2nd harmonic minus 3rd harmonic at time of <i>f0</i> maximum
h1minusa1p0_[V1,V2,V3]	1st harmonic minus amplitude of first formant at time of <i>f0</i> maximum
h1minusa2p0_[V1,V2,V3]	1st harmonic minus amplitude of second formant at time of <i>f0</i> maximum
h1minusa3p0_[V1,V2,V3]	1st harmonic minus amplitude of third formant at time of <i>f0</i> maximum
h1minush2F1_[V1,V2,V3]	1st harmonic minus 2nd harmonic at time of f1 maximum
h1minush3F1_[V1,V2,V3]	1st harmonic minus 3rd harmonic at time of f1 maximum
h2minush3F1_[V1,V2,V3]	2nd harmonic minus 3rd harmonic at time of f1 maximum

h1minusa1F1 [V1,V2,V3]

h1minusa2F1 [V1,V2,V3]

h1minusa3F1 [V1,V2,V3]

1st harmonic minus amplitude of first formant at time of f1 maximum

1st harmonic minus amplitude of second formant at time of f1 maximum

1st harmonic minus amplitude of third formant at time of f1 maximum

APPENDIX B

STIMULI FOR PERCEPTION EXPERIMENT

Item	Occurrence	Contrast condition	Sentence type	Focus category	Text
1	FOF/SOF	Degree modifier	declarative	s	At first, you made a very small amount more than I did. Then after a year or two you made much more than I did.
2	FOF/SOF	Degree modifier	declarative	s	I think Tom said it a little better than I did.
3	FOF/SOF	Degree modifier	declarative	ns	In fact, he said it a lot better than I did. I'll feel probably 90% better than I did last week. In fact, maybe 100% better than I did.
4	FOF/SOF	Degree modifier	declarative	ns	Today, I know a little bit more than I did when I started. And in a few weeks, I'll know way more than I did.
5	FOF/SOF	Adjective	declarative	s	You worked harder than I did, and you worked longer than I did.
6	FOF/SOF	Adjective	declarative	s	Tom knew more than I did, and he remembered more than I did.
7	FOF/SOF	Adjective	declarative	ns	I feel generally more pessimistic now than I did as a kid, and I feel more conservative than I did as a kid, as well.
8	FOF/SOF	Adjective	declarative	ns	I felt more comfortable onstage than I did offstage. And I felt more confident onstage than I did offstage.
9	FOF	--	declarative	s	There were a lot of photographers who would shoot more than I did.
10	FOF	--	declarative	s	He saw the situation differently than I did.
11	FOF	--	declarative	ns	I learned more in the last three hours than I did in the last three years of highschool.
12	FOF	--	declarative	ns	I've been traveling more than I did when I was playing full time, so it's time to slow down.
13	FOF	--	interrogative	s	Why would anyone stay there longer than I did?
14	FOF	--	interrogative	s	How can I help my kids to achieve more than I did?
15	FOF	--	interrogative	ns	Why do I have more energy today than I did the day before?
16	FOF	--	interrogative	ns	How can I find time to visit my family this year more than I did last year?

REFERENCES

- Abe, Shigeo. 2005. *Support Vector Machines for Pattern Classification*. London: Springer-Verlag.
- Ahn, Byron. 2008. The prosody of emphatic reflexives in English. Ms. *Presented at Experimental and Theoretical Advances in Prosody, Cornell University*, April 11-13.
- Ahdesmäki, M. and K. Strimmer. 2010. Feature selection in omics prediction problems using cat scores and false nondiscovery rate control. *Annals of Applied Statistics* 4:503-519.
- Aoun, J., and Y. A Li. 1993. *Syntax of scope*. Cambridge: MIT Press.
- Baayen, R. H., Davidson, D. J. and Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59:390-412.
- Baker, C. L. 1995. Contrast, discourse prominence and intensification, with special reference to locally free reflexives in British English. *Language* 71:63-101.
- Bard, Ellen, Daniel Robertson and Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72:32-68.
- Bartels, Christine. 2004. Acoustic correlates of 'second occurrence' focus: towards an experimental investigation. H. Kamp & B. Partee (eds.), *Context-dependence in the Analysis of Linguistic Meaning*, 354-61. Amsterdam: Elsevier.
- Bates, Douglas, Martin Maechler & Ben Bolker. 2009. *lme4*. R package available at <http://cran.r-project.org/web/packages/lme4/index.html>.
- Bayer, Joseph. 1996. *Directionality and Logical Form: On the Scope of Focusing Particles and Wh-in-situ*. Dordrecht; Boston; London: Kluwer.
- Beaver, David I. and Brady Clark. 2008. *Sense and Sensitivity: How Focus Determines Meaning*. Malden, MA: Blackwell.
- Beaver, David I. and Brady Z. Clark. 2003. "Always" and "Only": Why not all focus sensitive operators are alike. *Natural Language Semantics* 11:323-362.
- Beaver, David I., Brady Z. Clark, Flemming, Edward, Jaeger, Florian T., and Wolters, Maria. 2007. When semantics meets phonetics: acoustical studies of second occurrence focus. *Language* 83:245-76.
- Beaver, David I. and Dan Velleman. In press. The communicative significance of primary and secondary accents. *Lingua*.
- Becker, Joseph D. The phrasal lexicon. *Proceedings of the 1975 Workshop on Theoretical Issues in Natural Language Processing*, 40-63. Association for Computational Linguistics.
- Beckman, Mary E. 1996. The parsing of prosody. *Language and Cognitive Processes* 11:17-67.
- Beckman, Mary E. 1986. *Stress and Non-stress accent*. Dordrecht: Foris.
- Beckman, Mary E. Jennifer J. Venditti. 2010. Tone and intonation, in W.J. Hardcastle, J. Laver, and F.E. Gibbon (eds.), *The Handbook of Phonetic Sciences*, 603-652. Blackwell.
- Beckman, Mary E. and Gayle M. Ayers. 1994. *Guidelines for ToBI Labeling, vers. 2.0*. Ms. And accompanying speech materials. Ohio State University.

- Beckman, Mary E. and Jan Edwards. 1994. Articulatory evidence for differentiating stress categories. In P. Keating (ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology 3*, 7-33. Cambridge University Press.
- Beckman, Mary, Julia Hirschberg, and Stephanie Shattuck-Hufnagel. 2005. The original ToBI system and the evolution of the ToBI framework. *Prosodic typology: The phonology of intonation and phrasing* 9-54.
- Beckman, Mary E. and Janet B. Pierrehumbert. 1986. Intonational structure in English and Japanese. *Phonology Yearbook* 3:255-310.
- Bell, Alan and Jason Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky. 2009. Predictability Effects on Durations of Content and Function Words in Conversational English. *Journal of Memory and Language* 60:1, 92-111.
- Bergeton, Uffe. 2004. The Independence of Binding and Intensification: Volume 1. Doctoral dissertation: University of Southern California.
- Berners-Lee, Timothy. 1989. *Information Management: A Proposal*. CERN.
- Biber, Douglas and Edward Finegan. 1991. On the exploitation of computerized corpora in variation studies. In K. Aijmer and B. Altenberg (eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, 204-220. London: Longman.
- Bickerton, Derek. 1987. He himself: anaphor, pronoun, or ... ? *Linguistic Inquiry* 18:345-348.
- Bishop, Jason B. 2008. The effect of position on the realization of second occurrence focus. *Proceedings of Interspeech 2008*:869-872. International Speech Communication Association.
- Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5(9-10): 341-345.
- Bolinger, Dwight. 1981. *Two Kinds of Vowels, Two Kinds of Rhythm*. Reproduced by the Indiana University Linguistics Club.
- Bolinger, Dwight. 1961. Contrastive accent and contrastive stress. *Language* 37:83-96.
- Bolinger, Dwight. 1958a. A theory of pitch accent in English. *Word* 14:109-149.
- Bolinger, Dwight. 1958b. Stress and information. *American Speech* 33:5-12.
- Borsley, Robert D. 2005 Introduction. *Lingua* 115: 1475-1480.
- Bosch, Peter. 1988. Representing and accessing focussed referents. *Language and Cognitive Processes* 3:207-232.
- Boser, B. E, I. M Guyon, and V. N Vapnik. 1992. A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144-152.
- Brenier, Jason. 2008. *The Automatic Prediction of Prosodic Prominence from Text*. Doctoral dissertation: University of Colorado.
- Bresnan, Joan. 2006. *Why Laboratory Syntax?* Class notes. Stanford University. [<http://www.stanford.edu/~bresnan/LabSyntax2006/>]
- Brown, Michael T., & Tinsley, H. E. A. 1983. Discriminant analysis. *Journal of Leisure Research* 15, 290-310.
- Brown, Michael T. and Lori R. Wicker. 2000. Discriminant Analysis. In H. Tinsley *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, 209-235. San Diego: Academic Press.

- Browning, Marguerite A. Adverbial reflexives. *North East Linguistic Society (NELS)* 23:83-94.
- Bruce, Gösta. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerup.
- Büring, Daniel. 2008. Been there, marked that – a tentative theory of second occurrence focus. Ms. UCLA.
- Calhoun, Sasha. 2008. Why do we accent words? The processing of focus and prosodic Structure. Presented at *Experimental and Theoretical Advances in Prosody Conference*, Cornell University, NY.
- Calhoun, Sasha. 2007. Predicting focus through prominence structure. In *Proceedings of Interspeech 2007*, Antwerp, Belgium.
- Calhoun, Sasha. 2006. *Information Structure and the Prosodic Structure of English: a Probabilistic Relationship*. Doctoral dissertation: University of Edinburgh.
- Calhoun, Sasha, Malvina Nissim, Mark Steedman and Jason Brenier. 2005. A framework for annotating information structure in discourse. In *Frontiers in Corpus Annotation II: Pie in the Sky*, ACL2005 Conference Workshop, Ann Arbor, MI.
- Campbell, Nick and Mary Beckman. 1997. Stress, prominence, and spectral tilt. In G. Carayannis, A. Botinis and G. Kouroupetroglou (eds.), *Intonation: Theory, Models and Applications (Proceedings of an ESCA Workshop)*, September 18-20, 1997, Athens, Greece). ESCA and University of Athens Department of Informatics.
- Canavan, Alexandra, David Graff and George Zipperlen. 1997. *CALLHOME American English Speech*. Philadelphia: Linguistic Data Consortium.
- Cantrall, William R. 1973. Why I would relate ‘own’, emphatic reflexives, and intensive pronouns, my own self.’ In C. Corum, T.C. Smith-Stark & A. Weiser. *Papers from the Ninth Regional Meeting*, 57-67. Chicago Linguistic Society.
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and points of view. In C. Li (ed.), *Subject and Topic*, 25-56. New York: Academic Press.
- Chang, Chih and Chih Lin. 2001. LIBSVM: a library for support vector machines. URL <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Cho, Taehong. 2006. Manifestation of prosodic structure in articulation: Evidence from lip kinematics in English. In L. M. Goldstein, D.H. Whalen and C.T. Best (eds.), *Laboratory phonology 8: Varieties of phonological competence*, 519-548. Berlin/New York: Mouton de Gruyter.
- Cho, Taehong. 2005. Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /α,i/ in English. *Journal of the Acoustical Society of America* 117:3867-3878.
- Chomsky, Noam. 1972. *Language and Mind*. New York: Harcourt Brace Jovanovich.
- Chomsky, Noam. 1971. Deep structure, surface structure and semantic interpretation. In D. Steinberg and L. Jakobovits (eds.), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*. Cambridge: Cambridge University Press.
- Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.
- Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford: Oxford University Press.

- Cinque, Guglielmo. 1993. A null theory of phrase and compound stress. *Linguistic Inquiry* 24:239-297.
- Clark, H. H., and Haviland, S. E. 1977. Comprehension and the given–new contract. In R. O. Freedle (ed.), *Discourse Production and Comprehension*, 1–40. Hillsdale, NJ: Erlbaum.
- Clark, H. H., and Haviland, S. E. 1974. Psychological processes in linguistic explanation. In D. Cohen (ed.), *Explaining Linguistic Phenomena*: 91–124. Washington: Hemisphere Publication Corporation.
- Cohen, Dana. 1999. Towards a unified account of intensive reflexives. *Journal of Pragmatics* 31:1041-1052.
- Cohn, Abby, Cecile Fougeron, and Marie Huffman (eds.). In press. *Handbook of Laboratory Phonology*. Oxford, UK: Oxford University Press.
- Cole, Jennifer. 2011. Commentary on Harvesting Speech Datasets for Linguistic Research on the Web. Presentation at *Digging into Data Challenge Conference*. Washington, D.C. June 9-10.
- Cole, Jennifer and Mark Hasegawa-Johnson. In Press. Corpus phonology with speech resources. In A. Cohn, C. Fougeron, and M. Huffman (eds.), *Handbook of Laboratory Phonology*, Oxford, UK: Oxford University Press.
- Cooper, William, Stephen J. Eady and Pamela R. Mueller. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America* 77:2142-2155.
- Cortes, Corina and Vladimir Vapnik. 1995. Support vector networks. *Machine Learning* 20:273-297.
- Cowart, Wayne. 1997. *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks, CA: Sage.
- Creswell, Cassandre. 2002. The use of emphatic reflexives with NPs in English. In K. van Deemter & R. Kibble (eds.), *Information Sharing*, 31-62. Stanford, CA: CSLI Publications.
- Cristianini, Nello, and John Shawe-Taylor. 2000. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press.
- Cruttenden, Alan. 1986. *Intonation*. Cambridge, U.K.: Cambridge University Press.
- Crystal, David. 1969. *Prosodic Systems and Intonation in English*. Cambridge University Press.
- Cunningham, Emma. 2009. Adnominal ‘self’-intensification. Ms. Brown University.
- Dahl, Deborah and Jeanette K. Gundel. 1981. Identifying referents for two kinds of pronouns. *Minnesota Papers in Linguistics and Philosophy of Language* 8:10-29.
- Dainora, Audra. 2001. *An Empirically Based Probabilistic Model on Intonation in American English*. Doctoral dissertation: University of Chicago.
- Derwing, Bruce L. Psycholinguistic evidence and linguistic theory. In G. D. Prideaux (ed.), *Perspectives in experimental linguistics : papers from the University of Alberta Conference on Experimental Linguistics (Edmonton, 13-14 October 1978)*. Amsterdam: John Benjamins.
- Diaz-Uriarte, Ramon. 2009. *VarSelRF: Variable Selection Using Random Forests*. URL <http://ligarto.org/rdiaz/Software/Software.html>, R package version 0.7-1.

- Dilley, Laura, Mara Breen, Edward Gibson, M. Bolivar, and John Kraemer. 2006. A comparison of inter-transcriber reliability for two systems of prosodic annotation: RaP (rhythm and pitch) and ToBI (tones and break indices). In *The proceedings of the International Conference on Spoken Language Processing*, Pittsburg, PA.
- Dimitriadou, Evgenia, Kurt Hornik, Friedrich Leisch, David Meyer and Andreas Weingessel. 2009. *e1071: Misc Functions of the Department of Statistics*, TU Wien . R package version 1.5-19.
- D'Imperio, Mariapaola. 2002. Italian intonation: An overview and some questions. *Probus* 14, 37-69.
- Dirven, René. 1973. Emphatic and reflexive in English and Dutch. *Leuvense Bijdragen* 63:288-299.
- Dretske, Fred I. 1972. Contrastive statements. *The Philosophical Review* 81(4):411-437.
- Eady, S. J. and W. E. Cooper. 1986. Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America* 80: 402–415.
- Eady, Stephen J., William E. Cooper, Gayle V. Klouda, Pamela R. Mueller and Dan W. Lotts. 1986. Acoustical characteristics of sentential focus: narrow vs. broad and single vs. dual focus environments. *Language and Speech* 29:233-351.
- Eckardt, Regine. 2001. Reanalyzing 'selbst'. *Natural Language Semantics* 9.
- Edmondson, Jerold and Frans Plank. 1978. Great expectations: An intensive self analysis. *Linguistics and Philosophy* 2: 373-413.
- Evgeniou, T., M. Pontil, and T. Poggio. 2000. Regularization networks and support vector machines. *Advances in Computational Mathematics* 13:1–50.
- Face, Timothy L. 2002. Local intonational marking of Spanish contrastive focus. *Probus* 14, 71–92.
- Féry, Caroline and Shinichiro Ishihara. (2009). The phonology of second occurrence focus, *Journal of Linguistics* 45: 285–313.
- Féry, Caroline and Vieri Samek-Lodovici. 2006. Focus projection and prosodic prominence in nested foci. *Language* 82:131-150.
- von Fintel, Kai. 1994. *Restrictions on Quantifier Domains*. Doctoral dissertation: University of Massachusetts.
- Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7:179–188.
- Fletcher, William. 2007. Implementing a BNC Comparable Web Corpus. *Web as Corpus* 3.
- Fox, Anthony. 2000. *Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals*.
- Fowler, Carol A. 1995. Acoustic and kinematic correlates of contrastive stress in spoken English. In F. Bell-Berti & J.J. Raphael (eds.), *Producing Speech: Contemporary Issues: For Katherine Safford Harris*, 355-373.
- Fowler, Carol A. and Jonathan Housum. 1987. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26(5):489-504.

- Francis, W. Nelson and Henry Kucera. 1964. *A Standard Corpus of Present Day Edited American English, for Use with Digital Computers*. Providence, Rhode Island: Department of Linguistics Brown University.
- Friedman, Jerome H. 1989. Regularized discriminant analysis. *Journal of the American Statistical Association* 84: 165–175.
- Frota, Sonia. 2000. *Prosody and Focus in European Portuguese: Phonological Phrasing and Intonation*. New York, London: Garland.
- Garrod, S.C. and Sanford 1982. The mental representation of discourse in a focussed memory system: Implications for the interpretation of anaphoric noun phrases. *Journal of Semantics* 1:21-41.
- Gast, Volker. 2006. *The Grammar of Identity: Intensifiers and Reflexives in Germanic Languages*. Routledge: Florence, KY, USA.
- Gast, Volker. 2004. The interpretation of logophoric self-forms, and some consequences for a model of reference and denotation. In A. Branco, T. McEnery & R. Mitkov (eds.), *Proceedings of the 5th Discourse Anaphora and Anaphora Resolution Colloquium DAARC*, 75-80. Lisbon: Edições Colibri.
- Giegerich, Heinz J. *Metrical Phonology and Phonological Structure: German and English*. Cambridge: Cambridge University Press.
- Gilquin, Gaëtanelle and Stefan Th. Gries. Corpora and experimental methods: a state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5:1-26.
- Godfrey, John J. and Edward Holliman. 1997. *Switchboard-1 Release 2*, Linguistic Data Consortium, Philadelphia.
- Godfrey, John J., Edward Holliman and Jane McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *IEEE ICASSP-92. ACL Workshop on Discourse Annotation*.
- Golde, Karin. 1999. Evidence for two types of English intensive NPs. *Papers from the 35th Regional Meeting*. Chicago: Chicago Linguistics Society.
- Goldsmith, John. 1976. *Autosegmental Phonology*. Doctoral dissertation: MIT.
- Grice, H. Paul. 1975. Logic and conversation. In Peter Cole and Jerry Morgan (eds.), *Syntax and Semantics, Volume 3: Speech Acts*, 43–58. New York: Academic Press.
- Grosz, Barbara J. & Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12:175-204.
- Gundel, Jeanette K. 1999. On different kinds of focus. In P. Bosch and R. van der Sandt (eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives*, 293-305. Cambridge: Cambridge University Press.
- Gundel, Jeanette K., Nancy Hedberg and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69:274-307.
- Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Gussenhoven, Carlos. 1983. Focus, mode and the nucleus. *Journal of Linguistics* 19:377-417.
- Hajičová, Eva. 1984. Presupposition and allegation revisited. *Journal of Pragmatics* 8: 155-167.
- Hajičová, Eva. 1983. Topic and focus. *Theoretical Linguistics* 10:268-276.

- Halle, Morris and Jean-Roger Vergnaud. 1987. *An Essay on Stress*. Cambridge, M.A.: MIT Press.
- Halliday, M. A. K. 1967. *Intonation and Grammar in British English*. Mouton.
- Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. University of Chicago.
- Hayes, Bruce. 1981. *A Metrical Theory of Stress Rules*. Doctoral dissertation: MIT.
- Hayes, Bruce and Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language and Linguistic Theory* 9:47-96.
- HCRC Map Task Corpus. 1993. Philadelphia: Linguistic Data Consortium.
- Hirschberg, Julia and Gregory Ward. 1985. Implicating uncertainty: the pragmatics of fall-rise intonation. *Language* 61:747-776.
- Hockett, Charles F. 1958. *A Course in Modern Linguistics*. Macmillan.
- Hole, Daniel. 2002. Agentive selbst in German. In: Katz, G., Reinhard, S. & Reuter, P., eds. *Sinn und Bedeutung 6. Proceedings of the sixth meeting of the Gesellschaft für Semantik*, 133-150. Osnabrück: Institute of Cognitive Sciences.
- Horn, Larry. 1996. Exclusive company: Only and the dynamics of vertical inference. *Journal of Semantics* 13:1-40.
- Howell, Jonathan. 2010. Second occurrence focus and the acoustics of prominence. In *Interfaces in Linguistics: New Research Perspectives*, eds. R. Folli & C. Ulbrich. Oxford University Press, 278-298.
- Howell, Jonathan and Mats Rooth. 2009. Web harvest of minimal intonational pairs. In I. Alegria, I. Leturia & S. Sharoff (eds.), *Proceedings of the Fifth Web as Corpus Workshop*, 45-52. San Sebastian, Spain: Elhuyar Fundazioa.
- Huberty, Carl J. 2006. *Applied MANOVA and Discriminant Analysis*. 2nd ed. Hoboken, N.J: Wiley-Interscience.
- Huddleston, R. D., G. K. Pullum, and L. Bauer. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press New York.
- Hsu, C. W, C. C Chang, and C. J Lin. 2003. *A Practical Guide to Support Vector Classification*. Ms.
- Hyman, Larry. 1978. Tone and/or accent. In D. J. Napoli (ed.), *Elements of Tone, Stress and Intonation*, 1-20. Georgetown University Press.
- Idsardi, William J. 1992. *The Computation of Prosody*. Doctoral dissertation: MIT.
- Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar*. MIT Press.
- Jakobson, Roman. 1931/1962. Die Betonung und ihre Rolle in der Wort- und Syntagmaphonologie. In R. Jakobson & S. Rudy (eds.) *Selected Writings*. 'S-Gravenhage: Mouton 117-136.
- Jakobson, Roman, Gunnar Fant and Morris Halle. 1952. *Preliminaries to speech analysis*. Cambridge, M.A: MIT Press.
- Jannedy, Stefanie. *Hat Patterns and Double Peaks: The Phonetics and Psycholinguistics of Broad versus Late Narrow versus Double Focus Intonations*. 2002. Doctoral dissertation.
- Joachims, Thorsten. A probabilistic analysis of the rocchio algorithm with TFIDF for text categorization. *Proceedings of International Conference on Machine Learning (ICML)*, 1997.

- de Jong, Kenneth J. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America* 97:491–504.
- de Jong, Kenneth J. 1991. *The Oral Articulation of English Stress Accent*. Doctoral dissertation: Ohio State University.
- Jurafsky, Daniel, Alan Bell, Eric Fosler-Lussier, Cynthia Girand, and William Raymond. 1998. Reduction of English function words in Switchboard. *Proceedings of ICSLP-98*, Volume 7.
- Kadmon, Nirit. 2001. *Formal Pragmatics: Semantics, Pragmatics, Presupposition and Focus*. Malden: Blackwell.
- Keerthi, S. S., and C. J. Lin. 2003. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation* 15:1667–1689.
- Keller, Frank. 1998. Gradient grammaticality as an effect of selective constraint re-ranking. In M. C. Gruber, D. Higgins, K. S. Olson and T. Wysocki (eds.), *Papers from the 34th Meeting of the Chicago Linguistic Society*. Vol. 2: The Panels, 95–109. Chicago.
- Kemmer, Suzanne. 1995. Emphatic and reflexive -self: expectations, viewpoint, and subjectivity. In D. Stein and S. Wright (eds.), *Subjectivity and Subjectivism*, 55–82. Cambridge: Cambridge University Press.
- Kennedy, Christopher & Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 81:345–381.
- Kepser, Stefan and Marga Rice. 2005. Evidence in linguistics. In S. Kepser and M. Rice (eds.), *Linguistic Evidence : Empirical, Theoretical, and Computational Perspectives*, 1–6. Berlin: Walter de Gruyter.
- Keshet, Joseph, David Grangier, and Samy Bengio. 2009. Discriminative keyword spotting. *Speech Communication* 51(4):317–329.
- Kilgariff, Adam. 2007. Googleology is bad science. *Computational Linguistics* 33(1):147–151.
- Kilgariff, Adam & Gregory Grefenstette. 2003. Introduction to the special issue on the web as corpus. *Computational Linguistics* 29(3):333–347.
- Kingston, John and Mary Beckman. 1990. *Between the Grammar and Physics of Speech. Papers in Laboratory Phonology 1* (eds.). Cambridge; New York: Cambridge University Press.
- Kiss, Katlin É. 1998. Identificational focus vs. information focus. *Language* 74:245–273.
- Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59:1208–1221.
- Klecka, William. 1980. *Discriminant analysis*. Beverly Hills: Sage Publications.
- Klinedinst, Nathan. 2004. Only scalar only. *Paris Presupposition and Implicature Workshop*. 5 October.
- Kochanski, Gregory. 2006. Prosody beyond fundamental frequency. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzy I. Mleinek, N. Richter and J. Schließer (eds.), *Methods in Empirical Prosody Research*, 89–122. Berlin; New York: De Gruyter.
- Kohler, Joachim, Martha Larson, Franciska de Jong & Wesse Kraaij. 2008. Spoken content retrieval: searching spontaneous conversational speech. *Searching Spontaneous Conversational Speech 2008*.

- König, Ekkehard. 1991. *The Meaning of Focus Particles: A Comparative Perspective*, London: Routledge.
- König, Ekkehard and Volker Gast. 2006. Focused assertion of identity: A typology of intensifiers. *Linguistic Typology* 10:223-276.
- König, Ekkehard & Peter Siemund. 2000a. The development of complex reflexives and intensifiers in English. *Diachronica* XVII:39-84.
- König, Ekkehard & Peter Siemund. 2000b. Intensifiers and reflexives—a typological perspective. In Z. Frajzyngier & T. Curl (eds.) *Reflexives—Forms and Functions*, 41-74. Amsterdam; Philadelphia: Benjamins.
- König, Ekkehard & Peter Siemund. 2000c. Locally free *self*-forms, logophoricity and intensification. *English Language and Linguistics* 4:183-204.
- König, Ekkehard & Peter Siemund. 1996. Emphatische reflexiva und fokussstruktur', in I. Rosengren (ed.), *Sprache und Pragmatik* 40, Lund: Lunds Universitet, 1-42.
- Koopmans-van Beinum, F. J. and D. R. van Bergem. 1989. The role of 'given' and 'new' in the production and perception of vowel contrasts in read text and in spontaneous speech. *Proceedings of Eurospeech 1989*, 113-116.
- Kratzer, Angelika. 2004. Interpreting focus: presupposed or expressive meanings? *Theoretical Linguistics* 30:123-136.
- Kratzer, Angelika. 1991. The representation of focus. In A. von Stechow & D. Wunderlich (eds.), *Handbuch Semantik/Handbook Semantics*, 825-834. Berlin, New York: de Gruyter.
- Kratzer, Angelika and Elisabeth Selkirk. 2010. Distinguishing contrastive, new and given information. Presentation at the *Meeting of Semanticists Active in Canada (MOSAIC)*, June 1.
- Kratzer, Angelika, and Selkirk, Elisabeth. 2007. Phase theory and prosodic spellout: The case of verbs. *The Linguistic Review* 24, 93-135.
- Krifka, Manfred. 2004. Focus and/or context: a second look at second occurrence focus. In H. Kamp and B. Partee (eds.), *Context-Dependence in the Analysis of Linguistic Meaning*, 187-207. Amsterdam: Elsevier.
- Krifka, Manfred. 1992. A compositional semantics for multiple focus constructions. In S. Moore and A.Z. Wyner (eds.) *Proceedings of Semantics and Linguistic Theory 1*. Cornell University: CLC Publications.
- Kruttenden, Alan. 1997. *Intonation*. 2nd ed. Cambridge, U.K.: Cambridge University Press.
- Lachenbruch, Peter. 1975. *Discriminant analysis*. New York: Hafner Press.
- Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.
- Ladd, D. Robert. 1993. In defense of a metrical theory of intonational downstep. In H. van der Hulst & K. Snider (eds.) *The Phonology of Tone: The Representation of Ronal Register*, 109-132. Mouton de Gruyter.
- Ladd, D. Robert. 1991. One word's strength is another word's weakness: integrating syntagmatic and paradigmatic aspects of stress. *Proceedings of the Seventh Eastern States Conference on Linguistics*.
- Ladd, D. Robert. 1980. *The Structure of Intonational Meaning*. Bloomington: University of Indiana.
- Ladefoged, Peter. 1975. *A Course in Phonetics*. Orlando: Harcourt Brace.

- Ladefoged, Peter. 1967. *Three Areas of Experimental Phonetics*. London: Oxford University Press.
- Ladefoged, Peter and Loeb, Gerald. 2002. Preliminary studies on respiratory activity in speech. *UCLA Working Papers in Phonetics* 101:50-60.
- Leben, William. 1973. *Suprasegmental Phonology*. Doctoral dissertation: MIT.
- LeGac, David. 2002. Tonal alternations and prosodic structure in Somali. In B. Bel & I. Marlien, *Speech Prosody 2002: Proceedings of the 1st International Conference on Speech Prosody*. Aix-en-Provence.
- Legate, Julie Anne. 2001. Some interface properties of the phase. *Linguistic Inquiry* 34: 506-516.
- Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, M.A.: MIT Press.
- Leskosky, Richard J. 1972. Intensive reflexives. *Studies in the Linguistic Sciences* 2:42-65.
- Lieberman, Mark. 2006. Podzinger rejects Jesus. Language Log blog posting, January 25. URL <http://itre.cis.upenn.edu/~myl/languageblog/archives/002785.html>
- Lieberman, Mark. 1975. *The Intonational System of English*. Doctoral Dissertation: MIT.
- Lieberman, Mark and Janet Pierrehumbert. 1984. Intonational variance under changes in pitch range and length. In M. Aronoff and R. Oehrle (eds.), *Language sound structure*, 157-233. Cambridge, M.A.: MIT Press.
- Lieberman, Mark and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8:249-336.
- Lin, H.-T. and C.-J. Lin. 2003. *A Study on Sigmoid Kernels for SVM and the Training of Non-PSD Kernels by SMO-type Methods*. Technical report, Department of Computer Science, National Taiwan University.
- MacKay, Thomas J. 1991. He himself: undiscovering an anaphor. *Linguistic Inquiry* 22:368-373.
- Marcus, Mitchell P., Beatrice Santorini and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19:313-330.
- Materna, Pavel, Eva Hajičová and Petr Sgall. 1987. Redundant answers and topic-focus articulation. *Linguistics and Philosophy* 10:101-113.
- Meyer, D., F. Leisch, and K. Hornik. 2003. The support vector machine under test. *Neurocomputing* 55:169-186.
- Mihalcea, Rada & Moldovan, Dan. 1999. A method for word sense disambiguation of unrestricted text. *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, 152-158.
- Mo, Yoonsook. 2010. *Prosody Production and Perception with Conversational Speech*. Doctoral Dissertation: University of Illinois at Urbana-Champaign.
- Mukherjee, S., P. Tamayo, D. Slonim, A. Verri, T. Golub, J. Mesirov and T. Poggio. 1999. Support vector machine classification of microarray data. Center for Biological & Computational Learning (CBCL) Paper 182.
- Neeleman, Ad, and Kriszta Szendroi. 2004. Superman sentences. *Linguistic Inquiry* 35:149-159.
- Nenkova, Ani and Dan Jurafsky. 2007. Automatic detection of contrastive elements in spontaneous speech. In *Proceedings of the IEEE workshop on Automatic Speech Recognition and Understanding (ASRU)*, Kyoto, Japan.
- Nespor, Marina and Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris.

- Nespor, Marina and Irene Vogel. 1982. Prosodic domains of external sandhi rules. In H. van der Hulst and N. Smith (eds.), *The Structure of Phonological Representations: Part I*, 225–255. Dordrecht: Foris.
- Newman, Stanley. 1946. On the stress system of English. *Word* 2:171-187.
- O'Connor, J. D. and G. F. Arnold. 1973. *Intonation of colloquial English* (2nd edn). London: Longman.
- Ohala, John. 1975. Review of Lehiste 1970. *Language* 51: 736–40.
- Ostendorf, Mari, Patti Price and Stefanie Shattuck-Hufnagel. 1996. Boston University Radio Speech Corpus, Linguistic Data Consortium, Philadelphia.
- Palmer, Harold. 1922. *English Intonation, with Systematic Exercises*. Cambridge: Heffer.
- Partee, Barbara. 1999. Focus, quantification, and semantics-pragmatics issues. In P. Bosch and R. van der Sandt (eds.) *Focus: Linguistic, Cognitive, and Computational Perspectives*, 213-231. Cambridge University Press.
- Partee, Barbara. 1991. Topic, focus and quantification. *Semantics and Linguistic Theory 1*, 159-87.
- Peregrin, Jarda and Peter Sgall. 1986. An attempt at a framework for semantic interpretation of natural language. *Theoretical Linguistics* 13:37-73.
- Peterson, N. Reinholt. 1986. Perceptual compensation for segmentally conditioned fundamental frequency perturbation. *Phonetica* 43:31-42.
- Pierrehumbert, Janet. 1980. *The Phonology and Phonetics of English Intonation*. Doctoral Dissertation: MIT.
- Pierrehumbert and Beckman 1988. *Japanese Tone Structure*. Cambridge MA: MIT Press.
- Pierrehumbert, Janet & Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In P.R. Cohen, J. Morgan & M.E. Pollack (eds.) *Intentions in Communication*, 271-311.
- Pike, Kenneth L. 1945. *The Intonation of American English*. Ann Arbor: University of Michigan Press.
- Pitrelli, John, Mary E. Beckman and Julia Hirschberg. 1994. Evaluation of prosodic transcription labeling reliability. In *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Japan.
- Pitt, Mark, Laura Dilley, Keith Johnson, Scott Kiesling, William Raymond, Elizabeth Hume and Eric Fosler-Lussier. 2007. *Buckeye Corpus of Conversational Speech (2nd release)* [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).
- Primus, Beatrice. 1992. Selbst - Variants of a scalar adverb in German, in J. Jacobs (ed.), *Informationsstruktur und Grammatik*, 54–88. Opladen: Westdeutscher Verlag.
- Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14:19-100.
- Puchalla, Jochen. 2008. *Cutmp3*. <http://www.puchallaonline.de/cutmp3.html>.
- R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- Riester, Arndt. 2006. Only scalar. In J. Huitink & S. Katrenko (eds.), *Proceedings of the Eleventh ESSLLI Student Session*, 64-75.

- Resnik, Philip. 1999. Mining the web for bilingual text. *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, 527-534.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In L. Haegeman (ed.), *Elements of Grammar*. Dordrecht: Kluwer.
- Rochemont, Michael. 1986. *Focus in Generative Grammar*. Amsterdam and Philadelphia: J. Benjamins.
- van Rooij, Robert. 2002. *Relevance only*. In F. Bos, Foster & Matheson (eds), *Proceedings of the Sixth Workshop on the Semantics and Pragmatics of Dialogue (EDILOG 2002 4-6 September)*, 155-160. Edinburgh, UK.
- Rooth, Mats. 2009. Second occurrence focus and relativized stress F, in M. Zimmerman & C. Féry (eds.), *Information Structure: Theoretical, Typological, and Experimental Perspectives*. Oxford: Oxford University Press, 15–35.
- Rooth, Mats. 2008. Notions of focus anaphoricity. *Acta Linguistica* 55:277-285.
- Rooth, Mats. 1996a. On the interface principles for intonational focus. In T. Galloway and J. Spence (eds.) *Proceedings of Semantics and Linguistic Theory (SALT) VI*, 202-226, Ithaca, NY. Cornell University.
- Rooth, Mats. 1996b. *Focus*. In S. Lappin (ed.), *The Handbook of Contemporary Semantic Theory*, 271-297. London: Basil Blackwell.
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics*: 1:75-116.
- Rooth, Mats. 1985. *Association with Focus*. Doctoral Dissertation: University of Massachusetts.
- Ross, John Robert. 1967. *Constraints on Variables in Syntax*. Doctoral dissertation: MIT.
- Sampson, Geoffrey. 2001. *Empirical Linguistics*. London; New York: Continuum.
- Sarojini, Balakrishnan, Narayanasamy Ramaraj and Savarimuthu Nickolas. 2009. Enhancing the performance of LibSVM classifier by kernel f-score feature selection. In S. Ranka, S. Aluru, R. Buyya, Y.-C. Chung, S. Dua, A. Grama, S.K.S. Gupta, R. Kumar, & V.V. Phoha (eds.), *Contemporary Computing*, 533–543.
- Saussure, Ferdinand de. 1967[1916]. *Cours de Linguistique Générale*. Wiesbaden: Harrassowitz.
- Schütze, Carson. 2005. Thinking about what we are asking speakers to do. In S. Kepser and M. Rice (eds.), *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, 457-485. Berlin: Walter de Gruyter.
- Schütze, Carson. 1996. *The Empirical Base of Linguistics: Grammatical Judgments and Linguistic Methodology*. University of Chicago Press.
- Schwarzschild, Roger. 1999. GIVENness, Avoid F and other constraints on the placement of focus. *Natural Language Semantics* 7:141-177.
- Selkirk, Elisabeth O. 2008a. The relation between syntax and phonetics is indirect: the case of contrastive focus prosody. Presented at *Experimental and Theoretical Advances in Prosody*, Cornell University, April.
- Selkirk, Elisabeth O. 2008b. Contrastive focus, givenness and the unmarked status of "discourse-new". *Acta Linguistica Hungarica* 55:331-346.
- Selkirk, Elisabeth E. 2007. Focus prominence drives focus phrasing: Bengali intonation revisited. In C. Lee, M. Gordon & D. Büring (eds.), *Topic and Focus: A Cross-Linguistic Perspective*, 217-246. Dordrecht: Kluwer.

- Selkirk, Elisabeth. 2002. Contrastive focus vs. presentational focus: Prosodic evidence from right node raising in English. In B. Bel & I. Marlien (eds.), *Speech Prosody 2002: Proceedings of the 1st International Conference on Speech Prosody*, 643-646. Aix-en-Provence, France.
- Selkirk, Elisabeth O. 1996. The prosodic structure of function words. In J. L. Morgan & K. Demuth (eds.), *Signal to Syntax*. Mahwah, NJ: Lawrence Erlbaum. 187-213.
- Selkirk, Elisabeth O. 1995a. Sentence prosody: intonation, stress, and phrasing. In J.A. Goldsmith (ed.), *Handbook of Phonological Theory*. London: Blackwell.
- Selkirk, Elisabeth O. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, M.A.:MIT.
- Selkirk, Elisabeth O. 1980. The role of prosodic categories in English word stress. *Linguistic Inquiry* 11(3):563-605.
- Selkirk, Elisabeth. 1978. On prosodic structure and its relation to syntactic structure. In T. Fretheim (ed.), *Nordic prosody Vol. 2*, 111-140. Trondheim: TAPIR.
- Sgall, Peter. 1984. *Contributions to Functional Syntax, Semantics, and Language Comprehension*. Amsterdam: J. Benjamins.
- Sgall, Peter, Eva Hajičová and Jarmila Panevova. 1986. *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. Dordrecht: Reidel.
- Shields, Lynne W. and David Balota. 1991. Repetition and associative context effects in speech production. *Language and Speech* 34(1):47-55.
- Shriberg, Elisabeth, Rebecca Bates, Andreas Stolcke, Paul Taylor, Daniel Jurafsky, Klaus Ries, Noah Coccaro, Rachel Martin, Marie Meteer, and Carol Van Ess-Dykema. 1998. Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech* 41: 443-492.
- Siemund, Peter. 2000. *Intensifiers in English and German: A Comparison*. Routledge: NY.
- Silverman, Kim. 1986. F0 segmental cues depend on intonation: the case of the rise after voiced stops. *Phonetica* 43: 76-91.
- Silverman, Kim, Mary Beckman, John Pitrelli, Mari Ostendorf, Colin W. Wightman, Patti Price, Pierrehumbert, Janet & Julia Hirschberg. 1992. ToBI: A standard for labelling English prosody. *International Conference on Spoken Language Processing (ICSLP) 1992*.
- Sluijter, Agaath and Vincent van Heuven. 1996. Spectral balance as acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100:2471-2485.
- Sperber, Dan and Ira A. Noveck. 2004. Introduction. In D. Sperber & I. A. Noveck (eds.), *Experimental Pragmatics*, 1-24. New York: Palgrave Macmillan.
- Sridhar, Vivek Kumar Rangarajan, Ani Nenkova, Shrikanth Narayanan and Dan Jurafsky. 2008. Detecting prominence in conversational speech: pitch accent, givenness and focus. In *Proceedings of Speech Prosody*, Campinas, Brazil.
- von Stechow Arnim. 1991. Focusing and backgrounding operators, in A. Werner (ed.), *Discourse Particles: Descriptive and Theoretical Investigations on the Logical, Syntactic and Pragmatic Properties of Discourse Particles in German*. Amsterdam: John Benjamins, 37-84.
- von Stechow, Arnim. 1985/1989. *Focusing and Backgrounding Operators*. (Technical report.) Konstanz: University of Konstanz.

- Steedman, Mark. 2007. Information-structural semantics for English intonation. In C. Lee, M. Gordon & D. Büring (eds.), *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*, 245-264. Dordrecht: Springer.
- Steedman, Mark. 1991. Structure and intonation. *Language* 67:260-296.
- Stenberg, Daniel. 2008. *cURL and libcurl*, <http://curl.haxx.se>.
- Stevens, J. 2002. *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum.
- Syrdal, A.K. and J. McGory. 2000. Inter-transcriber reliability of ToBI prosodic labeling. In *The proceedings of the International Conference on Spoken Language Processing*, Beijing, China.
- Tancredi, Christopher. 1990. Syntactic association with focus. In *Proceedings from the First Meeting of the Formal Linguistic Society of Mid-America, University of Wisconsin at Madison*, 289–303.
- Tavano, Erin. 2006. *A Bound-variable Analysis of the Adverbial Emphatic Reflexive, or How I Wrote this Paper Myself*. Ms. University of Southern California.
- Terken, J. & S. G. Nooteboom. 1987. Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes* 2:145-163.
- Trager, George L. and H. L. Smith. 1951. *An Outline of English Structure*. Norman, OK: Battenburg Press.
- Trubetzkoy, Nikolai Sergeevič. 1939 Grundzüge der Phonologie. *Travaux du cercle linguistique de Prague* 7.
- Trubetzkoy, Nikolai Sergeevič. 1935. Die phonologischen Grenzsignale. *Proceedings of the 2nd International Congress of the Phonetic Sciences* 2: 45-49.
- Truckenbrodt, Hubert. 1995. *Phonological Phrases---Their Relation to Syntax, Focus, and Prominence*. Doctoral dissertation: MIT.
- Turk, Alice, Satsuki Nakai and Mariko Sugahara. 2006. Acoustic segment durations in prosodic research: a practical guide. In S. Sudhoff (ed.) *Methods in Empirical Prosody Research*.
- Vallduví, Enric and Maria Vilkuña. 1998. On rheme and kontrast. In *Syntax and Semantics* 79-108. New York: Academic Press.
- Vanderslice, Ralph and Peter Ladefoged. 1972. Binary suprasegmental features and transformational word-accentuation rules. *Language* 28:819-838.
- Venables, W. N, and B. D Ripley. 2002. *Modern Applied Statistics with S*. Springer verlag.
- Verheijen, C. R. 1983. *Reflexives and intensifiers in modern British English*. Rijksuniversiteit te Leiden.
- Wagner, Michael. 2010. *Contrastive Topics Decomposed*. Ms. McGill University.
- Wagner, Michael. 2006. Association by movement: evidence from NPI-licensing. *Natural Language Semantics* 14:297–324.
- Welby, Pauline. 2003. Effects of pitch accent position, type, and status on focus projection. *Language and Speech* 46:53-81.
- Wells, J. C. 1982. *Accents of English*. Cambridge, U.K.: Cambridge University Press.
- Weston, J. Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., & Vapnik, V. 2001. Feature selection for SVMs. *Advances in Neural Information Processing Systems*, 668–674. San Mateo, CA : Morgan Kaufmann.

- Yoon, Tae-Jin. 2007. *A Predictive Model of Prosody through Grammatical Interface: A Computational Approach*. Doctoral dissertation: University of Illinois, Urbana-Champaign.
- Yoon, Tae-Jin, Sandra Chavarría, Jennifer Cole and Mark Hasegawa-Johnson. 2004. Intertranscriber reliability of prosodic labeling on telephone conversation using ToBI. In *Proceedings of Interspeech*, 2722–2732, Jeju, Korea, 2004.
- Yuan, Jiahong and Mark Liberman. 2008. Speaker identification in the SCOUTUS corpus. *Journal of the Acoustical Society of America* 123:3878.
- Zubizarreta, Maria Luisa. 1998. *Prosody, Focus and Word Order*. Cambridge, Mass.: MIT Press.